



# Teoría de la Señal, Telemática y Comunicaciones:

Acercamiento a las ingenierías  
TIC a través de los trabajos fin  
de grado y master

Cursos 16/17, 17/18, 18/19, 19/20

Luz García Martínez  
Jorge Navarro Ortiz  
Editores



**Teoría de la Señal, Telemática y Comunicaciones:**

***Acercamiento a las ingenierías TIC a través de los trabajos fin de grado y máster***

***cursos 16/17, 17/18, 18/19, 19/20***

**Editores:**

**Luz García Martínez**

**Jorge Navarro Ortiz**



Departamento de  
Teoría de la Señal,  
Telemática y  
Comunicaciones

ISBN-13: 978-84-09-27541-0

Editores: Luz García Martínez y Jorge Navarro Ortiz (Dpto. Teoría de la Señal, Telemática y Comunicaciones de la Universidad de Granada)

El contenido de los trabajos que componen este libro es propiedad de los autores de los mismos y está protegido por los derechos que se recogen en la Ley de Propiedad Intelectual. Los autores autorizan la edición de este libro y su distribución, sin que esto, en ningún caso, implique una cesión a favor de la Universidad de Granada de cualesquiera derechos de propiedad intelectual sobre los contenidos de los trabajos. Ni la Universidad de Granada, ni los editores, serán responsables de aquellos actos que vulneren los derechos de propiedad intelectual sobre estos trabajos.

© 2021, los autores

Portada: Pilar Andrés Maldonado

Maquetación: Jorge Navarro Ortiz y Luz García Martínez

# Presentación

Los Trabajos Fin de Grado (TFG) y de Máster (TFM) representan un punto de inflexión en la vida de los estudiantes, una transición entre la vida académica y la vida profesional. Estos trabajos están orientados a la evaluación de competencias asociadas a sus títulos correspondientes, siendo necesarios para su desarrollo exitoso un alto grado de madurez, de conocimiento del contexto y de dominio de los contenidos relacionados con el título.

Debido a la importancia de estos trabajos, el *Departamento de Teoría de la Señal, Telemática y Comunicaciones* lleva ya convocadas ocho ediciones de los *Premios a los mejores TFG/TFM* (disponibles en <http://dtstc.ugr.es/premiospfc/>) tutorizados por sus profesores. Estos trabajos se han desarrollado fundamentalmente en el grado y el máster de Ingeniería de Telecomunicación, y en menor medida en los correspondientes de Ingeniería Informática.

Este libro incluye los 47 trabajos presentados en las ediciones V, VI, VII y VIII de los citados premios, correspondientes a los cursos 16/17, 17/18, 18/19 y 19/20. Entre las diversas temáticas destacan el *Internet de las cosas*, las *redes móviles e inalámbricas* y la *seguridad en red* en el área de *Ingeniería Telemática* y las *antenas*, el *procesado de señales biomédicas*, el *procesado de señales sísmicas* y el *procesado de voz* en el área de *Teoría de la Señal, Telemática y Comunicaciones*.

Estas tecnologías de la información y las comunicaciones (TIC) han cobrado en estos últimos meses, debido a la pandemia de la COVID-19, una innegable relevancia tanto para la sociedad actual como para todo tipo de sectores como la educación, la salud, la administración pública, las empresas, etcétera. Las TIC han posibilitado el teletrabajo, la teleeducación, el comercio electrónico, los servicios de entretenimiento *online*, las relaciones sociales *virtuales*, etcétera. Es indiscutible que la situación habría sido aún peor si las TIC no hubieran permitido el acercamiento en esta época de aislamiento forzoso. De hecho, los trabajos desarrollados durante el último curso 19/20 se han visto afectados por la pandemia, habiendo demostrado los estudiantes y sus tutores madurez, esfuerzo y capacidad de adaptación para lograr culminar con éxito estos trabajos.

Quiero terminar estas palabras expresando mi agradecimiento a los estudiantes y profesores por la labor realizada, así como mis mejores deseos para su nueva etapa profesional con el convencimiento de que contribuirán a mejorar la sociedad a la que nos debemos.

Un saludo,

Jorge

Jorge Navarro Ortiz

*Director del Dpto. de Teoría de la Señal, Telemática y Comunicaciones  
ETS Ingenierías Informática y de Telecomunicación  
Universidad Granada*



# Índice de contribuciones

## **Área de Ingeniería Telemática**

### **Análisis de tráfico**

- Diseño de características en Big Data para la monitorización estadística multivariante de red ..... 3  
*José Manuel García Giménez (tutor José Camacho Páez, curso 17/18)*
- Estimación de matrices de tráfico con mínimos cuadrados parciales y control estadístico multivariante ..... 9  
*Francisco José Cuberos Muñoz (tutor José Camacho Páez, curso 16/17)*
- Estimación de matrices de tráfico en series temporales con mínimos cuadrados parciales ..... 15  
*Irene Herrera López (tutor José Camacho Páez, curso 16/17)*

### **Aplicaciones relacionadas con la biomedicina**

- Aplicación de realidad virtual con control cefálico: SkyMet VR ..... 21  
*Gabriel Ávila Muñoz (tutor Miguel Ángel López Gordo, curso 19/20)*

### **Calidad de servicio**

- Análisis del tráfico Skype: Evaluación objetiva y subjetiva para diferentes condiciones de red ..... 27  
*César Senés Romo (tutor Jorge Navarro Ortiz, curso 16/17)*

### **Internet de las cosas**

- Implementación de un robot controlado desde un dispositivo móvil ..... 33  
*Javier Montijano Valdivia (tutores Juan José Ramos Muñoz, Jonathan Prados Garzón, curso 16/17)*
- LoRaWAN gateway and IoT low-cost mote prototype ..... 37  
*Ángel Guzmán Martínez (tutor Jorge Navarro Ortiz, curso 16/17)*
- Security improvement for LoRaWAN networks ..... 43  
*Natalia Chinchilla Romero (tutor Jorge Navarro Ortiz, curso 18/19)*
- Sistema de bajo coste basado en bobinas para la monitorización de la humedad del suelo en entornos agrícolas ..... 49  
*Pablo Rodríguez Martín (tutor Sandra Sendra Compte, curso 19/20)*
- Sistema de control para UAVs y reconocimiento de objetos ..... 55  
*Joaquín Lara Sarabia (tutores Pablo Ameigeiras Gutiérrez, Juan José Ramos Muñoz, curso 17/18)*

## **Herramientas web**

Herramienta para la importación de CVN a la plataforma SICA..... 61  
*Juan Carlos Angulo Santos (tutor Jorge Navarro Ortiz, curso 17/18)*

## **Redes móviles e inalámbricas**

Diseño e implementación de algoritmos de planificación para 5G ..... 67  
*Alejandro Ramírez Arroyo (tutores Juan Francisco Valenzuela Valdés, Pablo Padilla de la Torre, curso 18/19)*

Diseño y optimización de dispositivos de alta frecuencia para redes 5G ..... 71  
*Miguel Ángel Vaquero Blasco (tutores Juan Francisco Valenzuela Valdés, Pablo Padilla de la Torre, curso 17/18)*

High frequency components design for 5G communications ..... 77  
*Ángel Palomares Caballero (tutores Juan Francisco Valenzuela Valdés, Pablo Padilla de la Torre, curso 17/18)*

Integración del simulador de canal de propagación QuaDRiGa en un simulador de nivel de sistema ..... 83  
*Francisco Jesús Quero de Iabe Rosa (tutores Pablo Muñoz Luengo, Juan Francisco Valenzuela Valdés, curso 17/18)*

Interconexión de miembros de un equipo de emergencias en entornos sin cobertura usando dispositivos móviles ..... 87  
*Félix Delgado Ferro (tutor Jorge Navarro Ortiz, curso 19/20)*

Modeling and optimization for the cells switch off problem in ultradense 5G networks ..... 93  
*Pablo Helio Zapata Cano (tutores Pablo Muñoz Luengo, Juan Francisco Valenzuela Valdés, curso 18/19)*

Network slicing in 5G systems: key enablers, architectures, and performance evaluation ..... 99  
*José Antonio Ordóñez Lucena (tutores Pablo Ameigeiras Gutiérrez, Juan José Ramos Muñoz, curso 16/17)*

Radio resource allocation in private 5G networks ..... 105  
*Lorena Chinchilla Romero (tutores Pablo Muñoz Luengo, Pablo Ameigeiras Gutiérrez, curso 19/20)*

SDN and NFV in 5G networks: handover and automated network service scaling as use cases ..... 111  
*Óscar Adamuz Hinojosa (tutores Pablo Ameigeiras Gutiérrez, Juan José Ramos Muñoz, curso 16/17)*

Urban 5GRX: herramienta para el despliegue de 5G en entornos urbanos densos ..... 117  
*Francisco Jesús Quero de la Rosa (tutores Pablo Muñoz Luengo, Juan Francisco Valenzuela Valdés, curso 19/20)*

## **Seguridad en red**

- Aplicación móvil para detección y alerta de casos de cyberbullying ..... 123  
*Beatriz Quesada Tobar (tutor Pedro García Teodoro, curso 19/20)*
- Desarrollo de un crawler para análisis de la conectividad en la red I2P ..... 127  
*Alberto Abellán Galera (tutor Gabriel Maciá Fernández, curso 17/18)*
- Estudio y análisis de la conectividad en la Deep Web mediante el desarrollo de nuevas herramientas: c4i2p (crawling for I2P) en I2P ..... 133  
*Alberto Abellán Galera (tutores Gabriel Maciá Fernández, Roberto Magán Carrión, curso 19/20)*
- Implementación de un sniffer de información confidencial escrita en un navegador mediante técnicas de hooking ..... 139  
*José Miguel Noguera Comino (tutor Gabriel Maciá Fernández, curso 17/18)*
- Optimización de un sistema de detección de amenazas en capturas de tráfico de red mediante algoritmos evolutivos ..... 145  
*Pablo Merino Ávila (tutores Antonio M. Mora García, José Camacho Páez, curso 19/20)*

## **Servicios de red**

- Composición de cadenas de servicios con un algoritmo de optimización basado en Colonias de Hormigas ..... 151  
*Segundo Moreno Torres (tutor Antonio M. Mora García, curso 18/19)*

## **Videojuegos**

- Desarrollo de un videojuego e implementación de técnicas de mejora de la QoE . 157  
*Juan Miguel García Martínez (tutores Juan José Ramos Muñoz, Jonathan Prados Garzón, curso 17/18)*
- Diseño e implementación de un videojuego didáctico multijugador ..... 163  
*Pedro José Arias Perandrés (tutor Juan José Ramos Muñoz, curso 17/18)*
- eGamer.Zone ..... 169  
*Cristian Almohalla Gómez (tutor Juan José Ramos Muñoz, curso 17/18)*
- Juego de estrategia multiplataforma con almacenamiento en la nube: Cross General ..... 175  
*José Carreto Morales (tutor Juan José Ramos Muñoz, curso 17/18)*

# Área de Teoría de la Señal y Comunicaciones

## Antenas

Diseño de antena en array basada en tecnología SIW para milimétricas ..... 183  
*Bilal Hammu Mohamed (tutores Pablo Padilla de la Torre, Francisco Javier García Ruiz, curso 19/20)*

Diseño de antenas de bocina SIW para sistemas de medidas de antenas..... 189  
*Cleofás Segura Gómez (tutores Pablo Padilla de la Torre, Juan Francisco Valenzuela Valdés, curso 18/19)*

## Comunicaciones móviles

Prototipo de una estación base 4G usando Open Air Interface ..... 195  
*Francisco García Espigares (tutores Jorge Navarro Ortiz, José Carlos Segura Luna, curso 16/17)*

## Procesado de señales biomédicas

Análisis del registro espacial en imágenes de tomografía cerebral..... 201  
*Fernando Bayo Martínez (tutores Fermín Segovia Román, Diego Salas González, curso 19/20)*

Diagnóstico de la enfermedad de Alzheimer basado en el análisis de imágenes médicas cerebrales mediante redes neuronales ..... 207  
*Pablo Manresa Nebot (tutor Fermín Segovia Román, curso 19/20)*

Diseño e implementación de modelos que combinan diferentes modalidades de imagen médica cerebral para mejorar el diagnóstico asistido de la enfermedad de Alzheimer ..... 211  
*Sergio Ortiz Ruiz (tutor Fermín Segovia Román, curso 19/20)*

Extracción de información textural en neuroimagen..... 217  
*Miguel Bolívar Lupiáñez (tutores Diego Salas González, Javier Ramírez Pérez de Hinestrosa, curso 17/18)*

Segmentación de imágenes PET y SPECT usando campos aleatorios ocultos de Markov ..... 223  
*José Alberto Gómez López (tutor Fermín Segovia Román, curso 17/18)*

Segmentación del cuerpo estriado en imágenes FP-CIT SPECT para el diagnóstico del Parkinson ..... 229  
*José Miguel Mateos Ramos (tutor Diego Salas González, curso 19/20)*

Uso de líneas de contorno en imágenes [123I] FP-CIT SPECT para el diagnóstico del Parkinson ..... 235  
*Mario Molina Molina (tutores Diego Salas González, Javier Ramírez Pérez de Hinestrosa, Juan Manuel Górriz Sáez, curso 17/18)*

## **Procesado de señales sísmicas**

Uso de wavelets para reducción de ruido: aplicación en señales de voz y señales sísmicas ..... 241  
*Laura Gámiz Pérez (tutoras M<sup>a</sup> Carmen Benítez Ortúzar, Luz García Martínez, curso 16/17)*

## **Procesado de voz**

Deep neural networks based multichannel speech enhancement ..... 247  
*Juan Manuel Martín Doñas (tutor Ángel M. Gómez García, Antonio M. Peinado Herreros, curso 16/17)*

Desarrollo de un array ad-hoc de micrófonos basado en smartphones tipo Android ..... 253  
*Pablo Jesús García Martínez (tutores Antonio M. Peinado Herreros, Ángel M. Gómez García, curso 16/17)*

Desarrollo de un sistema de beamforming basado en redes neuronales profundas ..... 259  
*David Pareja Martos (tutores Ángel M. Gómez García, Antonio M. Peinado Herreros, curso 18/19)*

Detección de patologías degenerativas mediante voz ..... 265  
*Miriam González Atienza (tutores José Andrés González López, Antonio M. Peinado Herreros, curso 19/20)*

Técnicas de anti-spoofing robustas al ruido para verificación automática de locutores ..... 271  
*Alejandro Gómez Alanís (tutores Antonio M. Peinado Herreros, José Andrés González López, curso 17/18)*



# ***Ingeniería Telemática***



# DISEÑO DE CARACTERÍSTICAS EN BIG DATA PARA LA MOTORIZACIÓN ESTADÍSTICA MULTIVARIANTE DE RED

Autor: José Manuel García Giménez, e-mail: jgarciag@ugr.es

Tutor: José Camacho Páez, e-mail: josecamacho@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—El volumen de datos que circula por Internet ha crecido exponencialmente de la mano de las nuevas tecnologías. En el ámbito de la ciberseguridad, este fenómeno ha motivado la utilización de nuevas técnicas que permitan analizar este tipo de información de forma eficiente. En este contexto, la monitorización estadística multivariante de red (MSNM, del inglés, *multivariate statistical network monitoring*) es una técnica prometedora, ya que permite la detección de anomalías y proporciona información relevante para su diagnóstico. Este trabajo se fundamenta sobre el análisis de *Big Data* en el ámbito ciberseguridad basado en MSNM. Para este propósito, se han desarrollado dos herramientas: una para la extracción de características y otra para la diagnosis de anomalías de red detectadas con MSNM

**Palabras clave**—Análisis Multivariante, Big Data, Seguridad en Red, Detección de intrusos, Diagnóstico

## I. INTRODUCCIÓN

SEGÚN la *Data Breach Database*, en la primera mitad de 2017, hubo un total de 918 incidentes de robo de datos y, en promedio, 122 registros de datos fueron exfiltrados cada segundo, de los cuales solo un 4.6% estaban cifrados [1]. La cantidad de datos que se roban por ataque, así como la cantidad de ataques se ha visto incrementada [2].

La ciberseguridad se está haciendo cada vez más compleja porque las amenazas se están haciendo cada vez más avanzadas técnicamente. Nuevos tipos de ataques surgen continuamente y es muy difícil mantenerse al día para los sistemas y equipos de seguridad. Por otra parte, la introducción de las técnicas de *Big Data* en este contexto está provocando un gran cambio en el enfoque y las estrategias utilizadas para la prevención y detección de incidentes de seguridad [3].

Todos estos hechos han motivado el estudio de nuevas técnicas que permitan analizar grandes volúmenes de datos de manera eficiente con el fin de detectar intrusiones y anomalías de red. En este proyecto se va estudiar la monitorización estadística multivariante de red [4] (MSNM, del inglés, *Multivariate Statistical Network Monitoring*) que se basa en el uso de técnicas de análisis estadístico multivariante para la monitorización y gestión de seguridad en red. Utilizando esta metodología se puede obtener una buena capacidad de detección, además de información adicional sobre el evento detectado. De esta manera, se pretende reducir el tiempo desde que se produce el ataque a que se genera una respuesta, reduciendo así el impacto económico del propio ataque.

En este trabajo se presenta una herramienta, el FCParser[7], que sirve para procesar datos basados en texto para que su formato sea adecuado para su análisis. Esta herramienta tiene dos módulos, el parser y el deparser. El módulo principal, el parser, sirve para procesar datos. El segundo módulo, el deparser, sirve para simplificar el proceso de diagnosis de una anomalía.

## II. DATOS DERIVADOS DE LA RED

El volumen de datos con el que se trabaja en una red actual se ha visto incrementado de forma exponencial en los últimos años [5]. Este aumento en el volumen ha implicado también un vasto incremento en la cantidad de datos útiles para la monitorización de red.

Hay multitud de tipos de fuentes de datos y dependiendo del tipo de información, los datos se deben gestionar de manera distinta. En función del origen de la información, los datos de red se pueden clasificar en tres tipos: datos de registros e información de estado, datos de tráfico y datos de seguridad [6]. De la misma manera, dependiendo del formato de los datos, se pueden clasificar en dos tipos: datos estructurados y no estructurados.

### A. Clasificación de datos por formato

En el ámbito de la monitorización y la gestión de evento de red, los datos más habituales son los basados en texto o los que pueden ser vistos como tal. Por ejemplo: flujos de datos, registros de un cortafuegos, etc. Dependiendo de formato, este tipo de datos se pueden clasificar en datos estructurados y datos no estructurados.

Los datos estructurados son aquellos que tienen un alto nivel de orden o un formato fijo. Los campos que se presentan son siempre los mismos y suelen estar separados por un delimitador determinado. Por este motivo, la extracción de la información relevante es trivial, por ejemplo un archivo csv, en el que los campos están separados por comas.

Las fuentes de datos no estructuradas no siguen un formato fijo, aunque pueden presentar cierto grado de orden. Por ejemplo, un registro de errores de una aplicación, donde se muestra un texto con la información de cada error.

### B. Clasificación de datos por tipo de información

Según el tipo de información que contengan los datos de red relevantes en el contexto de la monitorización y gestión de eventos, se pueden clasificar en tres grandes grupos: datos de registro e información de estado, datos de tráfico y datos de seguridad.

Los datos de registro, también conocidos como *logs*, proporcionan información del estado de un sistema, del funcionamiento de las aplicaciones, así como del estado y funcionamiento de la propia red.

Los datos de tráfico engloban toda la información que se obtiene a partir de tráfico en la red. Hay datos desde muy bajo nivel (paquetes), hasta muy alto nivel (por ejemplo, estadísticas de protocolos).

Por último, los datos de seguridad incluyen todos los datos que provienen fuentes relacionadas directamente con los programas de gestión de la seguridad en red, como puede ser un cortafuegos o un antivirus. Este tipo de fuentes proporcionan datos de mucho interés porque ya han pasado un filtro inicial y se sabe que con un alto porcentaje de acierto, contienen información de peso para el analista que monitoriza o gestiona la seguridad de la red.

### III. ANÁLISIS DE DATOS

Aunque las herramientas que se han desarrollado en este proyecto son extensibles a multitud de técnicas de análisis, se origina en el seno de una metodología en 5 pasos con MSNM propuesta en [11]. Los 5 pasos de la metodología son: parsing, fusión, detección, diagnóstico y *deparsing*.

- 1) Parsing: este paso consiste en transformar los datos de red de cualquier índole en observaciones de características adecuadas para su análisis. En este caso se propone una estrategia de contadores acumulativos para contabilizar el número de veces que sucede algún evento determinado, este tipo de enfoque se ha utilizado anteriormente en análisis multivariante basando en componentes principales del que surge el MSNM [8].
- 2) Fusión: Este paso consiste en mezclar de forma adecuada los datos que se obtienen de todas las fuentes heterogéneas de red de las que se dispone. El tipo de *parsing* que se ha realizado facilita esta tarea porque homogeniza los los datos de las distintas fuentes. De esta manera, la fusión consiste en concatenar los datos de las distintas fuentes teniendo en cuenta el tiempo de muestreo de cada una.
- 3) Detección: Este paso consiste en determinar las observaciones que se consideran anómalas. Además, no solo detectar todos los eventos, sino que establecer una priorización entre los los mismos para determinar cuales son los más importantes. En este proyecto se ha utilizado una detección basada en MSNM. Esta técnica proporciona una buena tasa de detección y además, da información adicional útil para el diagnóstico de los incidentes.
- 4) Diagnóstico: El siguiente paso es el proceso de diagnóstico de los eventos que se han considerado anómalos. Para ello, se utiliza la herramienta oMEDA [13] [14]. Este algoritmo permite saber cuales son las características que afectan más a que una observación se considere

anómala. Esto proporciona una información relevante de cual puede ser el origen de los incidentes detectados.

- 5) *Deparsing*: El *deparsing* trata de llevar el proceso de diagnóstico un paso más allá. Este paso es dependiente de la técnicas de detección y diagnóstico que se utilicen. En este caso, que se utiliza MSNM y oMEDA, consiste en combinar la información de detección y diagnóstico para extraer las porciones de información original que más tienen que ver con la anomalía detectada, simplificando la tarea del analista. El analista vería una marca de tiempo en la cual se considera que hay una anomalía y una cantidad reducida de datos que debe analizar para encontrar el origen. De este modo, se consigue reducir el tiempo que transcurre entre que se produce una anomalía y se genera una respuesta.

### IV. PARSER

Los datos brutos obtenidos de las fuentes de datos de red no son adecuados para su análisis. Habitualmente, es necesario hacer un procesado previo para obtener una representación adecuada en matrices de características que recopilan la mayor cantidad de información relevante de los datos.

El módulo parser del FCParse es una herramienta ideada para efectuar el *parsing* de datos basados en texto de forma configurable para que se adapte a todo tipo de fuentes de datos. El programa utiliza distintos archivos de configuración para cada fuente para adaptarse a los múltiples formatos de los posibles datos.

En el contexto de la monitorización de red, una estrategia utilizada para la extracción de características es la utilización de contadores acumulativos [8]. Con el FCParse se propone utilizar una generalización de esta estrategia denominada *FaaC* (del inglés, *feature as a counter*). *FaaC* es un algoritmo que consiste en contar el número de veces que sucede algún evento, como por ejemplo la aparición de una palabra concreta en una entrada en registro. Con esta estrategia de pueden procesar todas las posibles fuentes de datos [9].

Basándose en esta estrategia, se propone un procedimiento en cinco pasos: muestreo temporal, extracción de *record*, agregación, transformación a observaciones y fusión de datos. En la Fig. 1 se muestra un diagrama de bloques donde quedan reflejados los cinco pasos.

El primer paso, el muestreo temporal, es un proceso opcional que consiste en dividir los datos de entrada en ventanas de tiempo de igual duración denominados intervalos de muestreo. A partir de cada intervalo se generará una observación.

La extracción de *records* es un paso intermedio para la metodología del *FaaC*. Un *record* se define como un grupo de variables extraídas de una entrada de datos. Las variables se definen como los distintos campos que contienen la información que se quiere utilizar para el conteo de los eventos, por ejemplo, puerto de entrada o protocolo. A continuación, se presenta un ejemplo de extracción de *record* a partir de un *log* que proviene de un sistema de detección de intrusos. Se puede observar que es una fuente de datos no estructurada. En la tabla I se recoge como sería una posible selección de variables para esa *log* en la columna izquierda y el *record* extraído en la columna derecha.

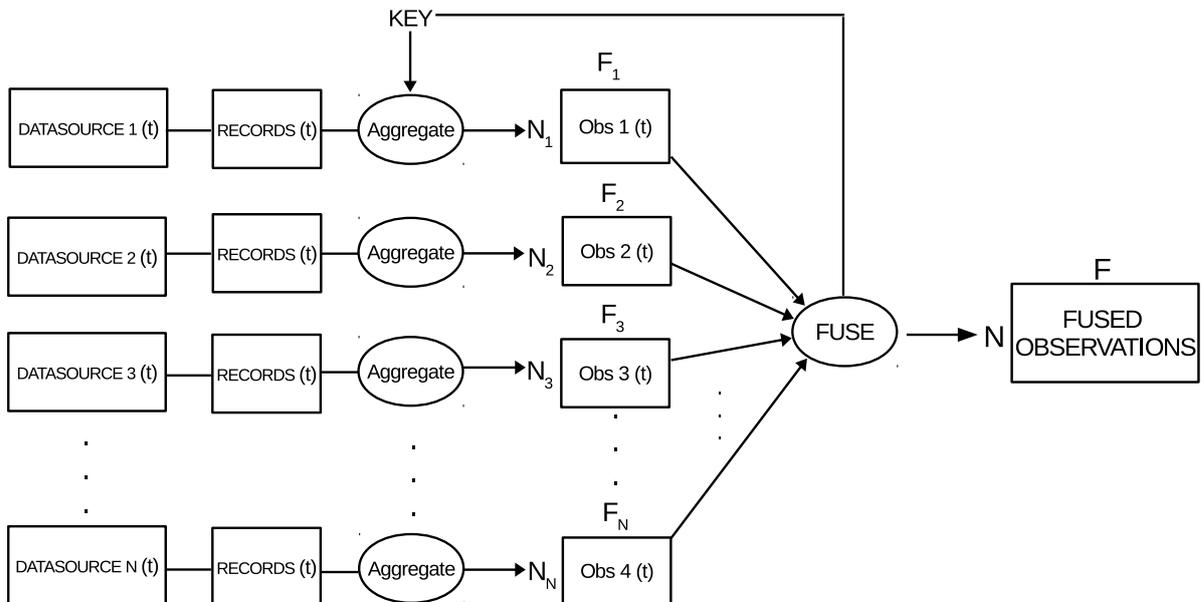


Fig. 1. Diagrama de bloques FCParser

Feb 27 00:56:43 bastion snort: [1:2000345:3] BLEEDING-EDGE IRC - Nick change on non-std port [Classification: A Network Trojan was detected] [Priority: 1]: TCP 11.11.79.67:1755 - 66.198.160.2:8888

El paso de agregación trata de agrupar los *records* según los distintos valores de una de las variables. La variable o variables que se eligen para este proceso se denominan *aggregation key*. Por ejemplo, si se elige la IP origen como *aggregation key*, se obtendrá una matriz de observaciones para cada IP distinta.

El cuarto paso es la conversión a observaciones. El programa está diseñado para generar una observación para cada intervalo de muestreo. Por esta razón, el FCParser necesita esperar a que la ventana temporal termine para conseguir una observación definitiva. Sin embargo, debido a la naturaleza de contadores aditivos de las características, las observaciones se pueden generar como la suma de las observaciones parciales del intervalo de muestreo. De esta manera, cuando se genera un *record*, este se transforma a una observación preliminar contando los eventos que se han detectado en ese intervalo de tiempo. Cuando la ventana de tiempo termine, se suman las observaciones preliminares para obtener la observación final.

El último paso es la fusión de datos. Este es el proceso

que integra los datos de las distintas fuentes de información para obtener el conocimiento relevante basado en las propias fuentes o la relación entre las mismas. En este caso, se va a utilizar la concatenación de las observaciones de las distintas fuentes porque los propios contadores homogeneizan las características. La concatenación consiste, simplemente, en unir los vectores de observaciones de un mismo intervalo de tiempo provenientes de distintas fuentes, dando lugar a una nueva observación.

#### A. Implementación

Primero, se ha creado una librería en python: Faaclib (del inglés, *feature as a counter library*). Esta librería contiene clases y funciones necesarias para los cinco pasos de la metodología. La librería puede funcionar como columna vertebral para cualquier herramienta basada en FaaC. La configuración del programa a cada fuente de datos se hace definiendo las variables y características que se pretenden extraer de dicha fuente. Tanto la librería, como la herramienta se encuentran disponibles en un repositorio público [7].

Las variables son los campos que se pueden extraer de una entrada de datos, se definen con tres parámetros: *name*, *matchtype* y *where*. *Name* es el nombre que el usuario le da a cada variable. *Where* es el parámetro que sirve para encontrar la variable dentro de una entrada de datos. En el caso de las fuentes estructuradas, es la posición del campo donde se encuentra la información. Por contrario, en las fuentes no estructuradas, es una expresión regular que sirve como patrón para encontrar el dato que se desea. En la *Tabla II* se puede ver un ejemplo de definición de variable con sus parámetros y unos posibles valores.

Por otro lado, las características, que son los contadores acumulativos de eventos, se pueden definir con 4 parámetros: *name*, *variable*, *matchtype* y *value*. *Name* es el nombre que se le quiera dar a la característica, *variable* es el campo que hay que mirar para ver si se cumplen las condiciones para

Tabla I  
EJEMPLO RECORD DE ENTRADA EN REGISTRO DE SNORT

VARIABLE	VALUE
Source ip	11.11.79.67
Protocol	TCP
Classification	A Network Trojan was detected
Destination IP	66.198.160.2
Label	BLEEDING-EDGE IRC - Nick change on non-std port
Priority	1
Destination port	8888
Timestamp	Feb 27 00:56:43
Source port	1755

que suceda el evento en cuestión. *Value* es el valor que debe tomar la variables para que se considere que el evento ha sucedido. Por último, *matchtype* es el tipo de dato del campo. En la *Tabla III* se muestra un ejemplo de definición de una características con los parámetros asociados.

Una vez se tienen los archivos de entrada finales, el procedimiento es el siguiente:

- 1) Primero se comienza a iterar por todos los archivos de cada fuente de datos.
- 2) Para cada archivo se va iterando por porciones de información necesarias para formar un *record*. Si la fuente es estructurada, se itera de línea en línea. Por contrario, si la fuente es no estructurada, se itera por los fragmentos de información que hay entre cada delimitador.
- 3) Con la información seleccionada se construye un *record*. Y éste se transforma en una observación preliminar.
- 4) Se procede al proceso de agregación si fuera necesario, dependiendo de la *aggregation key*.
- 5) Las observaciones preliminares se van acumulando en un *Observation batch*
- 6) Los *observation batch* a su vez se van agrupando en un diccionario en función de la fuente de la que provienen.
- 7) Por último se genera la salida del programa, formando los ficheros de salida con las observaciones.

## V. DEPARSER

Para este proyecto se parte de la base de que se dispone cierta información de las anomalías detectadas con el análisis multivariante, concretamente, la marca de tiempo en el que suceden las anomalías, así como, una lista de las características implicadas en cada una de ellas. Esa información se puede obtener en los pasos de detección y diagnóstico de la metodología en cinco pasos. El proceso de *deparsing* consiste en buscar los registros originales relacionados directamente con las anomalías detectadas.

Para intentar priorizar la información que se extrae de cada anomalía, se pretende que el programa dé como salida los registros asociados a la anomalías ordenados por cantidad de características implicadas que aparecen en cada registro.

Asimismo, para limitar la cantidad de información que pasa al analista, se propone la utilización de un umbral máximo de registros a extraer por cada fuente de datos. Sin embargo, no se pueden descartar registros que tienen la misma cantidad de

características referentes a anomalías ya que no hay manera de saber, a priori, si unos son más importantes que otros.

### A. Implementación

La herramienta está pensada para analizar anomalías de una en una, para que el analista pueda ir gestionando que anomalías analiza primero en función de la prioridad.

El procedimiento que se sigue en la implementación es el siguiente:

- 1) Se consigue la información referente a la anomalía y se extraen tanto las características implicadas, como la marca o marcas de tiempo en las que sucede.
- 2) Se buscan las fuentes de datos correspondientes a las características en cuestión.
- 3) Se hace una primera iteración buscando la cantidad de eventos relacionados con la anomalía suceden en cada registro. Para cada registro se comprueba si la marca de tiempo coincide con las que se están buscando. En caso afirmativo, se comprueba la cantidad de las características implicadas para la anomalía. Este proceso búsqueda es prácticamente igual el que había en el proceso de *parsing* para formar las observaciones y depende de si la fuente es estructurada o no estructurada.
- 4) Se comprueba el umbral. El algoritmo es el siguiente: se busca el bloque de *logs* que tenga el mayor número de características implicadas. Si el número de registros del bloque es mayor que el umbral, esos son los registros que dará como salida. Por contrario, si no se llega al umbral, se toma el siguiente bloque de una característica menos. Este proceso se repite hasta que se alcanza el umbral, o se acaban los registros.
- 5) Se genera el archivo de salida con los registros que cumplen las condiciones y tienen una cantidad de características por encima del umbral.

Por último, como salida, esta herramienta generará un archivo por cada una de las fuentes de datos de entrada. En cada uno de estos archivos se escribirán los registros de forma priorizada que tengan que ver, según las condiciones impuestas, con la anomalía que se quiera diagnosticar.

## VI. EXPERIMENTACIÓN

A continuación se va a evaluar la herramienta con un experimento. Para este fin, se ha utilizado un conjunto de datos propuesto para un desafío de *visual analytics* en el congreso *VisWeek*. El desafío viene de una de sus conferencias: el VAST del 2012(del inglés, *Visual Analytics Science and Technology*)[10]: *Bank of Money Regional Office Network Operations Forensics*. En este reto se presentan datos de seguridad de una red en explotación que ha tenido problemas de funcionamiento para conseguir arrojar algo de luz sobre lo que ha pasado en la red. Hay dos fuentes de datos, un log de cortafuegos y un log del sistema de detección de intrusos *Snort*.

Para ello, se va a utilizar la metodología en cinco pasos para el análisis de *Big Data* para la ciberseguridad. Para la implementación de la parte de análisis multivariantes, se ha utilizado la MEDAToolbox [12]. Todo esto, se ha implementado en una máquina virtual en la que se incluye tanto el

Tabla II  
EJEMPLO DE DEFINICIÓN DE VARIABLES

Parámetro	Valor
Name	operation
Matchtype	string
Where	'(?<=Priority: )(?P<match >[0-9]+)'

Tabla III  
EJEMPLO DE DEFINICIÓN DE CARACTERÍSTICAS

Parámetro	Valor
Name	ids_prio2
Variable	priority
Matchtype	single
Value	2

conjunto de datos, como las herramientas necesarias para el experimento. Junto a este material, se puede encontrar toda la documentación necesaria para reproducir las pruebas que se han utilizado para evaluar las herramientas desarrolladas.

### A. Parsing y fusión

De la primera fuente, el cortafuegos, se han seleccionado 142 características. En el caso de los datos del sistema de detección de intrusos, se ha extraído 163 características. En las tablas IV y V, se puede encontrar un resumen de las características seleccionadas. Se ha utilizado un muestreo de 1 minuto. Esto hace un total de 305 características y 2345 instantes de tiempo.

### B. Detección

Se va a realizar un análisis exploratorio de los datos con dos aproximaciones de análisis multivariante distintas: GPCA y PCA ya que no se dispone de un conjunto de calibración y uno de test. A partir de los resultados, se buscará la manera de priorizar e ir analizando los eventos más importantes.

1) *PCA*: En este caso, para decidir que eventos se deben estudiar en profundidad, se han utilizado percentiles de los estadísticos D-st y Q-st [9]. A partir del percentil, se establece un límite para analizar una cantidad reducida de eventos, que serán anómalos. Como estas observaciones tan anómalas afectan drásticamente al análisis estadístico, se procede a hacer una segunda iteración de este procedimiento eliminando las observaciones mencionadas anteriormente. Siguiendo este procedimiento, quedan un total de 10 eventos para el proceso de *deparsing*.

2) *GPCA*: La detección de anomalías con GPCA se basa en la monitorización de los gráficos de *scores* para las distintas PCs que se utilizan en el modelo [15]. En estas gráficas se pueden establecer límites de control y monitorizar las observaciones que sobrepasen dichos límites. Como en el caso anterior, las observaciones detectadas se eliminan del modelo, y se hace una segunda iteración. En total se han seleccionado 13 eventos anómalos para ser analizados con el proceso de *deparsing*.

Tabla IV  
CARACTERÍSTICAS FW

Cantidad	Tipo de característica
6	Tipos de IP origen y destino
103	Puertos origen y destino reservados o habituales
14	Mensajes ASA
7	Prioridad de syslog
2	Protocolo de transporte

Tabla V  
CARACTERÍSTICAS IDS

Cantidad	Tipo de característica
6	Tipos de IP origen y destino
103	Puertos origen y destino reservados o habituales
4	Prioridad Snort
34	Clasificación Snort
10	Descripción evento Snort
6	Información paquete IP

### C. Diagnosis

Para diagnosticar los eventos detectados, se le aplica a las observaciones el algoritmo oMEDA. De esta manera se obtiene para cada observación, una lista de las características que más influyen en cada uno de los incidentes.

A continuación, con los datos de la parte de detección y el resultado del algoritmo oMEDA, se realiza el *deparsing* con la herramienta desarrollada, para reducir la información necesaria para responder al incidente.

1) *PCA*: El *deparsing* se hace por separado para los diez eventos. A partir de la información obtenida se puede arrojar luz acerca de lo que ha pasado. En la tabla VI se recoge el número de *logs* que se obtienen, los *logs* únicos y las conclusiones que se derivan a partir de los mismos.

A partir de los eventos detectados, se puede obtener una información general bastante buena de lo que ha sucedido en la red. Se ven muchos tipos de intentos de exfiltración de datos y escaneos en busca de vulnerabilidades. Esto da indicios de que hay una botnet que ha infectado la red interna y está intentado exfiltrar información de la red al exterior.

En total, se han estudiado en profundidad tan solo 2414 *logs*, tan solo un **0.01%** de la cantidad total de *logs*. Además, de entre estos *logs*, solo se encuentran 22 tipos distintos, tan solo el **0.00009%**.

2) *GPCA*: A partir de los 13 eventos seleccionados, se realiza el proceso de *deparsing* para obtener la información reducida de cada evento. En la tabla VII se puede ver un resumen de los datos que se obtienen como resultado de este proceso.

A pesar de que GPCA da lugar a algún falso positivo, con esta metodología también se puede deducir que lo que ha ocurrido en la red. En este caso, tan solo un **0.09%** de los

Tabla VI  
PCA: *Deparsing* EVENTOS DETECTADOS

Time	N logs	Logs unicos	Info
05/04 18:07	58	1	Ataque servidor DNS
06/04 00:04	101	4	Exfiltración SSH y telnet
06/04 00:05	179	4	Exfiltración SNMP
06/04 00:24	80	1	Exfiltración VNC
06/04 00:33	171	1	Exfiltración FTP
06/04 01:13	19	4	Exfiltración SSH y telnet
06/04 17:27-28	1171	1	Ataque servidor DNS
06/04 17:44-45	11	4	Conexiones bloqueadas
06/04 18:30	622	1	Trafico FTP bloqueado
06/04 21:44-46	2	1	ARP poisoning

Tabla VII  
PCA: *Deparsing* EVENTOS DETECTADOS

Time	N logs	Logs unicos	Info
05/04 18:00-05	8	2	Inicio servicios
06/04 00:04	50	3	Exfiltración SSH y telnet
06/04 00:04-05	72	4	SNMP, SSH y telnet
06/04 00:05	53	3	Exfiltración SNMP
06/04 00:24	80	1	Conexiones puerto FTP
06/04 01:13; 03:08	18948	1	conexiones puerto 80
06/04 01:14	42	5	IRC malicioso
06/04 17:28	254	1	Ataque servidor DNS
06/04 17:44-45	11	6	Cambio configuración FW
06/04 18:07	3	1	IRC malicioso
06/04 18:44	14	1	Conexiones IRC
06/04 18:30	622	1	Exfiltración FTP
07/04 06:47	2004	6	Conexiones IRC

logs totales se ha utilizado para este fin, pero en términos de logs únicos, tan solo 35 logs, es decir un **0.00015%** del total.

## VII. CONCLUSIONES

Se ha desarrollado la herramienta de extracción de características para el análisis estadístico multivariante. Se ha conseguido que esta herramienta sea altamente configurable de manera que pueda servir para procesar cualquier tipo de información basada en texto, sin necesidad de modificar el programa, siguiendo la filosofía FaaC.

Asimismo, se ha desarrollado la herramienta para la diagnosis avanzada a partir de los datos obtenidos con el análisis estadístico multivariante. Con esta herramienta se consigue la cantidad de información y el tiempo de respuesta ante una anomalía.

La eficacia de las herramientas desarrolladas para este proyecto ha quedado patente con la experimentación basada en un conjunto de datos no estructurados de seguridad. Con la herramienta de extracción de característica se ha logrado transformar 4.4 GB de datos no estructurados en una matriz de 2345 x 305 sin perder información en el proceso. Esta información se ha utilizado para la detección de anomalías con dos metodologías de análisis exploratorio de datos distintas. De manera análoga, se ha comprobado que con la herramienta de diagnosis avanzada ha conseguido reducir la cantidad de información a analizar en un 99.9991% para la primera metodología y en un 99.9985% en el caso de la segunda. A partir de esta información de diagnosis se consigue inferir el problema que ha causado las eventualidades en la red.

## AGRADECIMIENTOS

Este trabajo está parcialmente financiado por el Ministerio de Economía y competitividad y los fondos FEDER a través de los proyectos TIN2014-60346-R y TIN2017-83494-R y la Universidad de Granada, la Junta de Andalucía y el Fondo Social Europeo por la contratación de joven personal investigador, en el marco del sistema nacional de garantía juvenil y del programa operativo de empleo juvenil.

## REFERENCIAS

[1] Gemalto, "Gemalto Breach Level Index Report," p. 16, 2017.  
 [2] Ponemon Institute LLC, "2017 Cost of Data Breach Study" March, pp. 1-34, 2017.  
 [3] R. Eastman, M. Versace, and A. Webber, "Big data and predictive analytics: On the cybersecurity front line," IDC Whitepaper, February, 2015.  
 [4] J. Camacho, A. Pérez-Villegas, P. García-Teodoro, and G. Maciá-Fernández, "PCA-based multivariate statistical network monitoring for anomaly detection," Computers and Security, vol. 59, pp. 118-137, jun 2016.  
 [5] Cisco Systems, "White paper: The Zettabyte Era: Trends and Analysis," tech. rep., Cisco, 2010.  
 [6] J. Camacho, "Lecture notes in Gestión de incidentes de seguridad" February 2018. Master propio ciberseguridad, Universidad de Granada.  
 [7] J. M. García-Giménez, A. Pérez-Villegas, and J. Camacho, "Fcparser." <https://github.com/josecamachop/FCParser>, 2017.  
 [8] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," ACM SIGCOMM Computer Communication Review, vol. 34, no. 4, p. 219, 2004.  
 [9] J. Camacho, A. Pérez-Villegas, P. García-Teodoro, and G. Maciá Fernández, "PCA-based multivariate statistical network monitoring for anomaly detection," Computers and Security, vol. 59, pp. 118-137, jun 2016.  
 [10] "Visual analytics community." Retrieved from: <http://www.vacomunity.org/VAST+Challenge+2017>.

[11] J. Camacho, J. M. García-Giménez, N.M. Fuentes-García, G. Maciá-Fernández, and P. García-Teodoro, "Multivariate Big Data Analysis for Cybersecurity: 5 steps from the haystack to the needle," 2018. Manuscript in preparation.  
 [12] J. Camacho, A. Pérez-Villegas, R. A. Rodríguez-Gómez, and E. Jiménez-Mañas, "Multivariate Exploratory Data Analysis (MEDA) Toolbox for Matlab," Chemometrics and Intelligent Laboratory Systems, vol. 143, pp. 49-57, apr 2015.  
 [13] J. Camacho, "Observation-based missing data methods for exploratory data analysis to unveil the connection between observations and variables in latent subspace models," Journal of Chemometrics, vol. 25, no. 11, pp. 592-600, 2011.  
 [14] M. Fuentes-García, G. Maciá-Fernández, and J. Camacho, "Evaluation of diagnosis methods in pca-based multivariate statistical process control," Chemometrics and Intelligent Laboratory Systems, vol. 172, pp. 194 - 210, 2018.  
 [15] J. Camacho, R. A. Rodríguez-Gómez, and E. Saccenti, "Group-Wise Principal Component Analysis for Exploratory Data Analysis," Journal of Computational and Graphical Statistics, vol. 26, no. 3, pp. 501-512, 2017.



**Jose Manuel García Giménez** Graduado en ingeniería de tecnologías de telecomunicación en 2016 por la universidad de granada. Actualmente cursando el Máster de ingeniería de telecomunicaciones en la misma universidad. Ha colaborado con el departamento de teoría de la señal, telemática y comunicaciones de la Universidad de Granada. Sus intereses se centran en en la detección de anomalías y seguridad de red, utilizando técnicas de análisis de datos.



**Jose Camacho Páez** es profesor asociado al departamento de Teoría de la Señal, Telemática y Comunicaciones en la Universidad de Granada. Obtuvo la licenciatura en Ingeniería de telecomunicaciones por la universidad de Granada en 2003 y el grado de Doctor por la universidad de Valencia en 2007. Sus intereses en la investigación se centran en el análisis exploratorio de datos, la detección de anomalías y la optimización con técnicas multivariantes aplicadas a todo tipo de datos, incluyendo procesos de fabricación, quimiometría y redes de comunicación. Está interesado especialmente en la utilización de técnicas de análisis exploratorio de datos en Big Data para la ciberseguridad.

# Estimación de Matrices de Tráfico con Mínimos Cuadrados Parciales y Control Estadístico Multivariante

Autor: Francisco José Cuberos Muñoz, e-mail: francicm92@correo.ugr.es

Tutor: José Camacho Páez, e-mail: josecamacho@ugr.es

Titulación: Máster Universitario en Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—El objetivo del proyecto es presentar una solución alternativa para resolver el problema de estimación de la Matriz de Tráfico (MT). Se introduce una técnica novedosa en este campo: Mínimos Cuadrados Parciales (PLS), método que permite crear un modelo predictivo para estimar la MT a partir del protocolo Simple Network Management Protocol (SNMP). Los modelos predictivos basados en datos funcionan mal en extrapolación, por lo que para evitar que PLS trabaje en extrapolación se empleará la metodología de Control Estadístico Multivariante (MSPC), también novedosa en este campo de aplicación.

El objetivo es detectar cuándo ocurre este suceso y de esta forma evitarlo mediante la sustitución de la estimación de PLS por la de General Tomogravity (GT), que no es un modelo predictivo. Se obtendrá así una matriz denominada Matriz de Tráfico Híbrida (MTH) formada por las estimaciones obtenidas mediante GT y PLS. Una vez calculada, podrá ser utilizada en numerosos campos de aplicación de gran importancia relacionados con el análisis de la red.

**Palabras clave**—Matriz de Tráfico, estimación, Control Estadístico Multivariante, Mínimos Cuadrados Parciales, Matriz de Tráfico Híbrida.

## I. INTRODUCCIÓN A LA ESTIMACIÓN DE MATRICES DE TRÁFICO

La Matriz de Tráfico (denominada por sus siglas, MT), como su nombre indica, mide el volumen total de tráfico que se transmite desde cualquier nodo de entrada (origen) a cualquier nodo de salida (destino), en una red de comunicaciones durante un determinado periodo de tiempo. Las unidades del tráfico medido pueden ser *bytes* o paquetes, entre otros. En el contexto de Internet, estos nodos pueden representar *routers*, puntos de presencia (*Points of Presence*, PoP), redes de área local (*Local Area Network*, LAN), etc. La MT posee un amplio abanico de aplicaciones, como por ejemplo la predicción de tendencias futuras en el tráfico de red, la optimización de red, el diseño de protocolos, el análisis de riesgos o la detección de anomalías.

## II. ESTIMACIÓN DE MATRICES DE TRÁFICO

Una amplia visión de la carga de tráfico en la red puede obtenerse gracias a la Matriz de Tráfico, pues contiene la cantidad de tráfico existente en cada par origen-destino (OD) de la red.

Se denomina  $\mathbf{x}_t$  a la vectorización de la MT en un instante de tiempo  $t$ . El valor de  $\mathbf{x}_t$  puede medirse directamente

haciendo uso de un protocolo de registro de flujos de tráfico conocido como *Netflow* [1]. Los sensores de *Netflow* se encuentran localizados en diferentes partes de la red donde el tráfico entrante es registrado, midiendo de esta forma uno o varios elementos de  $\mathbf{x}_t$ .

Sin embargo, *Netflow* requiere una alta demanda de recursos en términos de tráfico generado, tiempo de procesamiento y capacidad de almacenamiento en los *routers*.

Una alternativa es estimar la MT a partir de los contadores de tráfico de los enlaces haciendo uso del protocolo SNMP (*Simple Network Management Protocol*).

La relación entre los contadores de los enlaces y la MT en un instante  $t$  dado es expresada por la siguiente expresión:

$$\mathbf{y}_t = \mathbf{x}_t \mathbf{A}^T \quad (1)$$

donde  $\mathbf{y}_t$  es un vector de dimensión  $L \times 1$  que contiene los contadores de tráfico de los enlaces, siendo  $L$  el número de enlaces;  $\mathbf{x}_t$  es la MT a estimar organizada como un vector de dimensión  $I \times N$ , donde  $N$  es el número de pares OD; y  $\mathbf{A}$  es la matriz de *routing* cuya dimensión es  $L \times N$ .

Finalmente, realizar la estimación de  $\mathbf{x}_t$  a partir de  $\mathbf{y}_t$  es un problema complejo debido a que el número de pares OD es mucho mayor que el número de enlaces. Además, es una de las razones por las que medir los flujos requiere muchos más recursos que obtener las mediciones de tráfico de los enlaces.

### II-A. Modelo de gravedad

El modelo de Gravedad (en inglés *Gravity Modeling*) [2], [3] es una de las aproximaciones más simples para el cálculo de la MT. Este modelo expresa que el intercambio de información entre distintas localizaciones es proporcional al volumen de tráfico entrante y saliente en dichas localizaciones. A su vez, puede dividirse en dos modelos: *Simple Gravity* y *General Gravity*.

### II-B. Modelo General Tomogravity

El método *Tomogravity* [3], [4], [5] consiste en dos pasos básicos:

1. Se obtiene una estimación inicial de la MT aplicando el modelo de Gravedad (resolviendo el modelo GG o el modelo SG).

2. La estimación inicial es combinada con el método de tomografía. Se aplica programación cuadrática para encontrar una solución que minimiza la norma de Frobenius de la diferencia entre la matriz a estimar y la estimación obtenida inicialmente, sujeta a determinadas restricciones. El problema de programación cuadrática que plantea el método de tomografía es el siguiente:

$$\text{minimizar} \quad \|\mathbf{X} - \mathbf{X}_g\| \quad (2)$$

$$\text{sujeta a} \quad \|\mathbf{A}\mathbf{X} - \mathbf{Y}\| \quad (3)$$

donde  $\|\cdot\|$  es la norma de Frobenius de una matriz,  $\mathbf{X}$  es la solución a obtener,  $\mathbf{X}_g$  es la matriz de tráfico estimada en el paso inicial,  $\mathbf{A}$  es la matriz de *routing* e  $\mathbf{Y}$  contiene los contadores de tráfico de los enlaces.

### II-C. Análisis por Componentes Principales

El Análisis por Componentes Principales (en inglés Principal Component Analysis, PCA) [6], [7], [8], [9], es una técnica multivariante que consiste en encontrar el subespacio de máxima varianza en el espacio de variables (eje de coordenadas cartesianas) de  $M$  dimensiones, haciendo uso de una matriz para calibración  $\mathbf{X}$  (datos originales) de dimensiones  $N \times M$ . La matriz  $\mathbf{X}$  está compuesta por  $N$  objetos, cada uno de los cuales posee las observaciones de  $M$  variables.

PCA se encarga de reducir la dimensión de los datos medidos ( $\mathbf{X}$ ) a un subespacio de dimensión  $A$  PCs ( $A \leq \text{rango}(\mathbf{X})$ ), en el cual la variabilidad de los datos viene dada por los PCs.

PCA queda definido por la siguiente expresión:

$$\mathbf{X} = \underbrace{\mathbf{T} \cdot \mathbf{P}^T}_{\text{Parte sistemática}} + \underbrace{\mathbf{E}}_{\text{Parte residual}} \quad (4)$$

donde  $\mathbf{T}$  ( $N \times A$ ) es la matriz de *scores* que contiene la proyección ortogonal de las observaciones en el subespacio de dimensión  $A$ , cuyas filas son las proyecciones de los datos en dicho subespacio;  $\mathbf{P}$  ( $M \times A$ ) es la matriz de *loadings* que contiene los vectores de *loading*, siendo estos los autovectores correspondientes a los  $A$  autovalores más altos de la matriz de covarianzas ( $\mathbf{C} = \frac{1}{N-1} \cdot \mathbf{X}^T \cdot \mathbf{X}$ ) de los datos originales ( $\mathbf{X}$ ), y que define las direcciones de mayor variabilidad del nuevo subespacio  $A$  dimensional; y  $\mathbf{E}$  es la matriz que contiene los residuos, es decir, la información que no es detallada por el modelo PCA.

## III. MATERIALES Y MÉTODOS

### III-A. Red Abilene

La red Abilene (actualmente denominada Internet2) [10], [11] se encontraba compuesta en 2003 por 12 *routers*, distribuidos a lo largo de Estados Unidos (Figura 1). Según esta topología de red, existen 144 pares OD ( $N = 12^2 = 144$ ).



Figura 1: Topología red Abilene (2003).

### III-B. Métricas de rendimiento

Con el fin de evaluar el rendimiento de la metodología propuesta, se introducen algunas medidas para evaluar el error. La medida de rendimiento más utilizada en la estimación de MT es el *Error Relativo Medio*, que en este caso será dividido en *Error Relativo Temporal* y *Error Relativo Espacial*.

**III-B1. Error Relativo Temporal:** El Error Relativo Temporal (en inglés *Temporal Relative Error*, TRE) [5], [12] ofrece una medida del error en cada instante de tiempo realizando la suma de los errores de todos los flujos OD en ese instante. Se define mediante la expresión [12]:

$$TRE(t) = \frac{\|\mathbf{X}^*(t) - \mathbf{X}(t)\|_2}{\|\mathbf{X}(t)\|_2} \quad (5)$$

donde  $\mathbf{X}^*(t)$  es la MT estimada y  $\mathbf{X}(t)$  es la matriz real en una determinada muestra  $t$ .

Además, también se hará uso de la Función de Distribución Acumulada (en inglés *Cumulative Distribution Function*, CDF) del TRE como medida del error.

**III-B2. Error Relativo Espacial:** El Error Relativo Espacial (en inglés *Spatial Relative Error*, SRE) [5], [12] obtiene una medida del error medio por flujo OD durante un intervalo de tiempo. Se define por [12]:

$$SRE(k) = \frac{\|\mathbf{X}_T^*(k) - \mathbf{X}_T(k)\|_2}{\|\mathbf{X}_T(k)\|_2} \quad (6)$$

donde  $\mathbf{X}_T^*$  y  $\mathbf{X}_T$  son los elementos estimados y reales en el vector de tráfico del  $k$ -ésimo flujo OD durante el intervalo de tiempo estimado.

Al igual que ocurre en el TRE, en este caso también se hará uso de la CDF del SRE como medida del error.

## IV. MÍNIMOS CUADRADOS PARCIALES

### IV-A. Fundamento teórico

El método de Mínimos Cuadrados Parciales (en inglés *Partial Least Squares*, PLS) [13], [7], [14], [15] es una técnica de análisis multivariante que relaciona dos matrices de datos,  $\mathbf{X}$  e  $\mathbf{Y}$ . La decisión principal cuando se calibra el modelo PLS es identificar un número reducido de nuevas variables, denominadas Variables Latentes (en inglés *Latent Variables*, LVs). Las LVs en PLS es el equivalente a los PCs en PCA.

PLS presenta una solución al problema de regresión lineal mediante una solución basada en mínimos cuadrados. El problema de regresión es definido a continuación:

$$\mathbf{X} = \mathbf{Y} \cdot \mathbf{B} + \mathbf{F} \quad (7)$$

donde  $\mathbf{X}$  ( $N \times K$ ) es la matriz de variables a ser estimada,  $\mathbf{Y}$  ( $N \times M$ ) es la matriz de variables a partir de la que se va a estimar  $\mathbf{X}$ ,  $\mathbf{B}$  ( $M \times K$ ) es la matriz de coeficientes de regresión y  $\mathbf{F}$  ( $N \times K$ ) es la matriz de residuos. Una posible forma de interpretar  $\mathbf{B}$  es como modelo de  $\mathbf{X}$ , siendo  $\mathbf{Y}$  la entrada del modelo.

Para el problema de regresión lineal de la ecuación 7 se presenta una solución denominada de mínimos cuadrados:

$$\hat{\mathbf{B}}_{PLS} = (\mathbf{Y}^T \cdot \mathbf{Y})^{-1} \cdot \mathbf{Y}^T \cdot \mathbf{X} \quad (8)$$

La inversión de la matriz  $\mathbf{Y}^T \cdot \mathbf{Y}$  es no singular, es decir, debe ser invertible. Dicha solución es muy inestable debido a una estimación pobre cuando esta matriz está mal condicionada. El buen condicionamiento de una matriz depende del grado de independencia existente entre las variables. Por lo tanto, si las variables de  $\mathbf{Y}$  son altamente colineales<sup>1</sup>, la estimación de  $\mathbf{X}$  no puede realizarse directamente de  $\mathbf{Y}$ .

Como alternativa, se va a predecir  $\mathbf{X}$  a partir del subespacio formado por  $A$  LVs en  $\mathbf{Y}$ , el cual maximiza su covarianza con  $\mathbf{X}$ . El problema de regresión lineal parcial entre las matrices  $\mathbf{X}$  e  $\mathbf{Y}$  puede establecerse como sigue:

$$\mathbf{Y} = \mathbf{T} \cdot \mathbf{P}^T + \mathbf{E} \quad (9)$$

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{Q}^T + \mathbf{F} \quad (10)$$

donde  $\mathbf{T}$  ( $N \times A$ ) es la matriz de *scores* que contiene las proyecciones de  $\mathbf{Y}$  en el subespacio formado por  $A$  LVs;  $\mathbf{P}$  ( $M \times A$ ) y  $\mathbf{Q}$  ( $K \times A$ ) son las matrices de regresión, también llamadas matrices de *loadings*; y  $\mathbf{E}$  ( $N \times M$ ) y  $\mathbf{F}$  ( $N \times K$ ) son las matrices de residuos de  $\mathbf{Y}$  y  $\mathbf{X}$  respectivamente.

De esta forma, la ecuación 7 puede ser redefinida de la siguiente forma:

$$\mathbf{X} = \mathbf{Y} \cdot \mathbf{B} + \mathbf{F} \quad (11)$$

con

$$\mathbf{B} = \mathbf{R} \cdot \mathbf{Q}^T \quad (12)$$

Obteniendo:

$$\mathbf{X} = \mathbf{Y} \cdot \mathbf{R} \cdot \mathbf{Q}^T + \mathbf{F} \quad (13)$$

donde

$$\mathbf{R} = \mathbf{W} \cdot (\mathbf{P}^T \cdot \mathbf{W})^{-1} \quad (14)$$

Por lo tanto, la solución de mínimos cuadrados parciales obtenida es:

$$\hat{\mathbf{B}}_{PLS} = \mathbf{W} \cdot (\mathbf{P}^T \cdot \mathbf{W})^{-1} \cdot \mathbf{Q}^T \quad (15)$$

donde  $\mathbf{W}$  ( $M \times A$ ) es la matriz de pesos. Así, el modelo PLS queda representado por las matrices  $\mathbf{P}$ ,  $\mathbf{W}$  y  $\mathbf{Q}$ .

<sup>1</sup>La colinealidad se produce cuando una variable es combinación lineal de otra.

#### IV-B. Problema extrapolación

Tal y como se mencionó con anterioridad, se han introducido algunas métricas de rendimiento para evaluar el error medio de las diferentes metodologías y de esta forma determinar cuál de ellas es más precisa. A continuación se aplicará el *Error Relativo Temporal* (TRE), el cual ofrece una medida del error en cada instante de tiempo realizando la suma de los errores de todos los flujos OD en ese instante.

A continuación, en la Figura 2 se muestra el rendimiento de las diferentes metodologías evaluando el TRE:

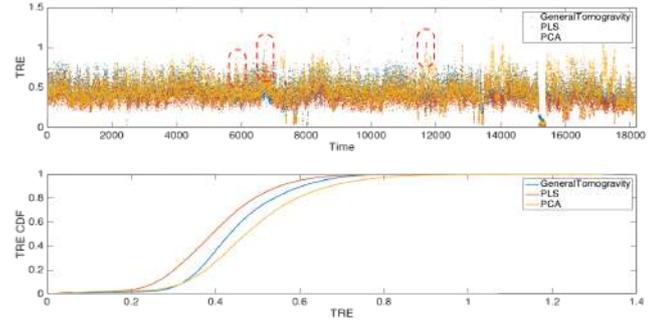


Figura 2: Rendimiento de las diferentes metodologías.

Puede observarse como la metodología propuesta basada en PLS tiene, de forma general, menor error que las demás. Sin embargo, PLS presenta un problema y es que existen observaciones cuyo error aumenta de forma notoria. Cuando se genera el subespacio formado por un número determinado de LVs a partir de un conjunto de datos de calibración, puede ocurrir que la observación de test (observación a predecir) se encuentre dentro de dicho subespacio y PLS pueda predecirla. O por el contrario, puede ocurrir que la observación se encuentre fuera de dicho subespacio y por lo tanto, en este caso, PLS no podría predecirla y trabajaría en extrapolación.

Cuando una observación se encuentra en el subespacio generado puede ser predicha con una precisión aceptable, sin embargo, debido a que las observaciones no tienen un comportamiento lineal, puede ocurrir que la precisión de predicción varíe según la observación en cuestión. Una vez detectadas las observaciones que extrapolan y por tanto no pueden ser predichas por PLS, es necesario sustituirlas por las observaciones correspondientes del modelo GT, siendo este un modelo no predictivo.

El problema de extrapolación se representa gráficamente en las Figuras 3 y 4.

#### V. CONTROL ESTADÍSTICO MULTIVARIANTE

##### V-A. D-estadístico

El D-estadístico (o estadístico *Hotelling's T<sup>2</sup>*) [16] [17] consiste en analizar la matriz de *scores*  $\mathbf{T}$  para comprobar la variabilidad de los datos proyectados en el nuevo subespacio formado por  $A$  Variables Latentes (en inglés *Latent Variables*, LVs). Representa la distancia de Mahalanobis estimada desde el centro del subespacio a la proyección de la observación dentro de este.

Para una observación  $n$ , el D-estadístico puede calcularse a partir de la siguiente expresión [6]:

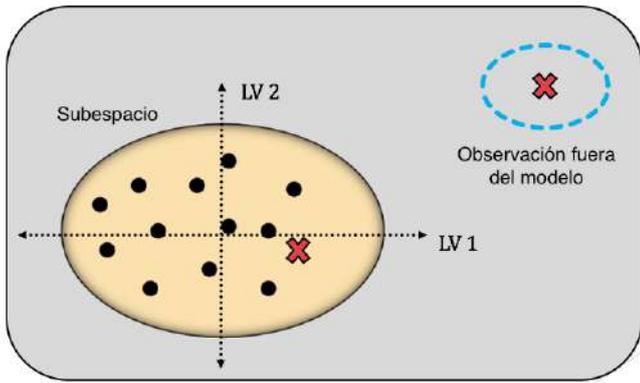


Figura 3: Extrapolación del modelo PLS (a).

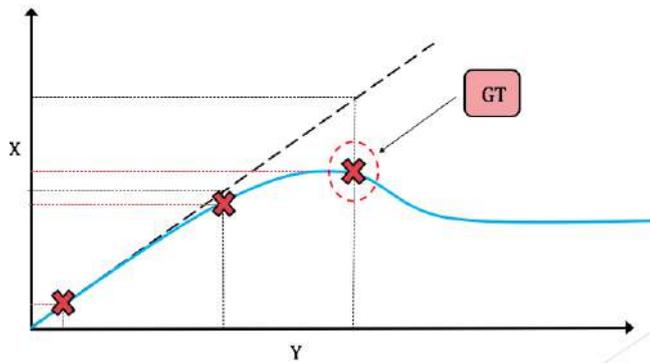


Figura 4: Extrapolación del modelo PLS (b).

$$D_n = \sum_{a=1}^A \left( \frac{t_{an} - \mu_{ta}}{\sigma_{ta}} \right)^2 \quad (16)$$

donde  $t_{an}$  representa el *score* de la observación en cuestión en la LV  $a$  y  $\mu_{ta}$  y  $\sigma_{ta}$  son la media y desviación estándar de los *scores* de esa LV en los datos de calibración.

Los *scores* son combinaciones lineales de las variables originales, por lo que de acuerdo con el Teorema Central del Límite, su función de distribución debe aproximarse a una distribución Normal (distribución de probabilidad de variable continua también denominada *distribución de Gauss*). Asumiendo esto, el D-estadístico sigue una distribución beta [6].

$$D \sim \frac{(N-1)^2}{N} B_{A/2, (N-A-1)/2} \quad (17)$$

Sin embargo, en el caso de disponer de un conjunto de datos grande, el D-estadístico puede aproximarse con una distribución F (distribución de probabilidad continua también conocida como *distribución F de Snedecor*) con  $A$  y  $N-A$  grados de libertad [6].

$$D \sim \frac{A \cdot (N^2 - 1)}{N \cdot (N - A)} F_{A, (N-A)} \quad (18)$$

Conociendo la distribución asociada, puede definirse un límite de control superior (en inglés *Upper Control Limit*, UCL) a un determinado nivel de confianza  $\alpha$ , cuya expresión viene dada por [6].

$$UCL(D)_\alpha = \frac{A \cdot (N^2 - 1)}{N \cdot (N - A)} F_{(A, (N-A)), \alpha} \quad (19)$$

Si el D-estadístico para una observación se encuentra fuera del límite de control, supondrá una variación inusual dentro el modelo. La distancia entre la observación en cuestión y el centro del modelo será grande. Es decir, existe una mayor distancia de Mahalanobis al centro del modelo, obtenido a partir de las observaciones que se encuentran en condiciones normales de operación. En otras palabras, es una medida de la "normalidad" de cada observación dentro del modelo.

### V-B. Q-estadístico

El Q-estadístico (o *Squared Prediction Errors*, SPE) [18] [16] [19] representa la distancia Euclídea de una observación al subespacio de  $A$  dimensiones, es decir, proporciona una medida de la proximidad de dicha observación al subespacio  $A$  dimensional. Una observación con un Q-estadístico muy grande supondrá una variación inusual fuera del modelo.

El cálculo del Q-estadístico para una observación  $n$  viene dado por [6]:

$$Q_n = \sum_{m=1}^M (e_{nm})^2 \quad (20)$$

donde  $e_{nm}$  es el valor residual correspondiente a la variable  $m$ .

Su correspondiente límite de control superior (UCL) [6] a un determinado nivel de confianza  $\alpha$  viene dado por:

$$UCL(Q)_\alpha = \theta_1 \cdot \left[ \frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \quad (21)$$

donde  $\theta_n = \sum_{a=A+1}^{rango(X)} (\lambda_a)^n$  con  $\lambda_a$  los autovalores de la matriz  $\frac{1}{N-1} \cdot \mathbf{E}^T \cdot \mathbf{E}$ , donde  $\mathbf{E}$  es la matriz residual;  $h_0 = 1 - \frac{2\theta_2 \theta_3}{3\theta_1^2}$  y  $z_\alpha$  es el percentil normal estandarizado a  $100 \cdot (1 - \alpha)\%$ .

Normalmente, el Q-estadístico posee mayor capacidad de detección que el D-estadístico [20].

Respecto al nivel de confianza  $\alpha$ , suele considerarse un valor de 0.01 o 0.05. En este caso, se utilizará  $\alpha = 0,01$  tanto para el D-estadístico como para el Q-estadístico.

En la Figura 5 puede observarse la representación gráfica de la definición tanto del Q-estadístico como del D-estadístico. En color rojo, se muestra un ejemplo de observación con un valor D alto, lo que significa que posee una mayor distancia de Mahalanobis al centro del subespacio formado por  $A$  LVs. Sin embargo, la observación se encuentra dentro del subespacio. Esto indica que se ha producido un comportamiento inusual dentro del modelo. Por otra parte, en azul se muestra un ejemplo de una observación con Q alto, lo que deriva en que la muestra se encuentra fuera del subespacio y por lo tanto se produce extrapolación del modelo. Es muy importante prestar atención a este último suceso.

De esta forma, la Figura 5 permite una mejor comprensión de la función de los estadísticos D y Q.

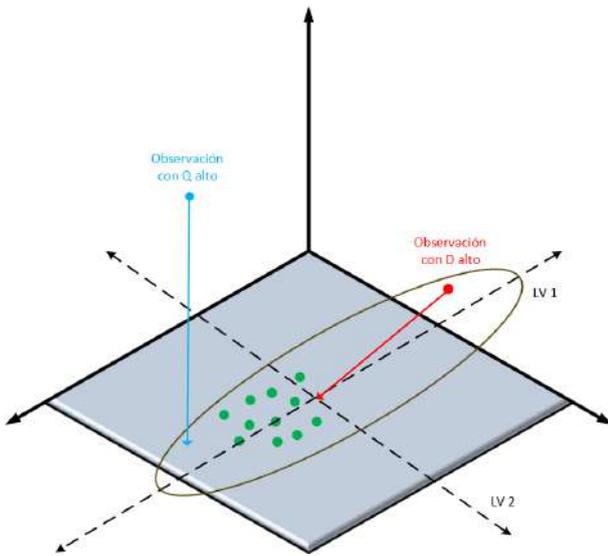


Figura 5: Interpretación de los estadísticos D y Q.

V-C. Aplicación de MSPC a PLS

En este apartado se aplicará la metodología MSPC a la metodología PLS implementada. En la Figura 6, además de representar el TRE, se incluyen los estadísticos D y Q aplicados a PLS. Tal y como se comentó con anterioridad, el Q-estadístico posee mayor capacidad de detección que el D-estadístico, esto es debido a que el Q-estadístico detecta cuándo se produce la extrapolación del modelo PLS, es decir, cuándo las observaciones se encuentran fuera del subespacio formado por A LVs. De esta forma, es importante prestar atención al Q-estadístico.

En la Figura 6 puede apreciarse cómo los instantes en los que se produce extrapolación del modelo PLS corresponden con las muestras de mayor TRE.

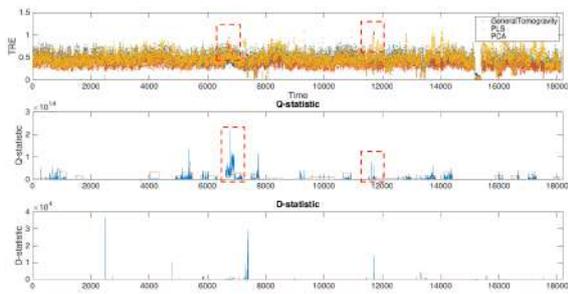


Figura 6: Aplicación de la metodología MSPC a PLS.

Sin embargo, es adecuado complementar el análisis del Q-estadístico con el D-estadístico.

El D-estadístico representa la distancia de Mahalanobis desde el centro del subespacio formado por A LVs hasta la proyección de la observación en cuestión. Por lo tanto, es posible que una observación tenga un D alto (distancia grande al centro del subespacio) y esto quede reflejado de forma clara en el gráfico de monitorización del D-estadístico, sin embargo, puesto que se encuentra dentro del subespacio puede ocurrir que no sea apreciable en el gráfico del Q-estadístico. De esta forma, ambos gráficos se complementan el uno al otro.

VI. RESULTADOS EXPERIMENTALES

Uno de los objetivos principales es evitar que PLS trabaje en extrapolación y para que esto no ocurra, dependiendo de los resultados ofrecidos por la metodología MSPC, se seleccionarán las observaciones correspondientes a GT o PLS dependiendo de si las observaciones superan los límites de control o no. De esta forma, se obtendrá la Matriz de Tráfico Híbrida (MTH) compuesta por observaciones de GT y PLS, es decir, se obtiene una matriz final formada por un método predictivo y otro no predictivo.

Para una mejor apreciación del resultado, en la Figura 7 se compara el TRE de la diferentes metodologías, donde se comprueba como la gráfica correspondiente a la MTH es la más próxima a cero, es decir, la de menor error. En la gráfica de la CDF, sin embargo, es difícil hacer una distinción entre PLS y MTH debido a la proximidad de sus respectivos valores.

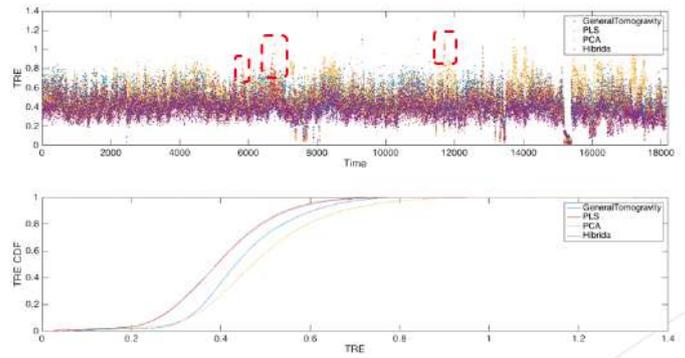


Figura 7: Comparación TRE final.

En la Figura 8 se hace lo propio con el SRE y en la Tabla I se muestran los valores finales del TRE y SRE para las diferentes metodologías. Se puede observar que la MTH calculada es la estimación que posee menor TRE, mejorando a GT y PCA así como también a PLS, el cual ya proporcionaba mejoría en los resultados. Sin embargo, la MTH presenta un valor de SRE más elevado que PLS y PCA. Esto es debido a que el SRE posee una debilidad. Si el tráfico actual de un flujo correspondiente a un par OD determinado es muy bajo, entonces el SRE puede ser alto. No es inusual que algunos pares OD experimenten pequeños periodos de tiempo donde el tráfico sea prácticamente nulo [21] debido a diversas razones.

Por lo tanto, si el tráfico de un par OD es muy cercano a cero, entonces el SRE puede ser demasiado grande.

VII. CONCLUSIONES

Como conclusión, se ha conseguido obtener mejoras de gran importancia en un problema de relevancia en la planificación de redes, el cual ha centrado numerosos proyectos de investigación en los últimos años debido a la enorme

Metodología	TRE	SRE
GT	0.4474	0.8456
PCA	0.4840	0.5747
PLS	0.3972	0.5164
<b>MTH</b>	<b>0.3955</b>	<b>0.5927</b>

Tabla I: Valores medios finales de TRE y SRE.

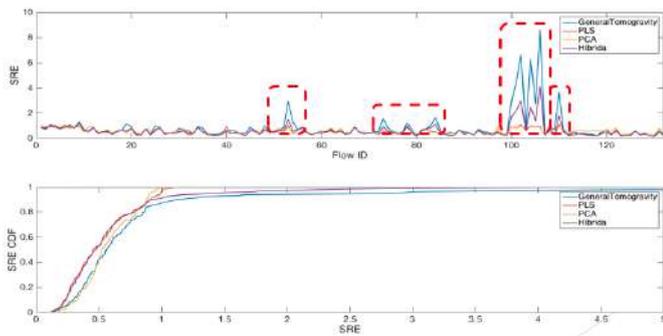


Figura 8: Comparación SRE final.

importancia que tiene realizar una buena estimación de la Matriz de Tráfico.

Se ha conseguido obtener una Matriz de Tráfico denominada Matriz de Tráfico Híbrida, compuesta tanto por un método predictivo como es PLS, como por un método no predictivo como General Tomography. Dicha matriz ha permitido corregir la extrapolación de PLS mejorando incluso los buenos resultados ofrecidos por este último. Adicionalmente, también han sido mejorados los resultados ofrecidos por las técnicas de estimación existentes hasta el momento.

La matriz obtenida podrá ser de gran utilidad en numerosas aplicaciones de ingeniería de redes. Como pueden ser por ejemplo, la predicción de la tendencia futura del tráfico, la optimización de la red, el diseño de protocolos o la detección de anomalías.

#### AGRADECIMIENTOS

Dar las gracias en primer lugar a Pepe, gran tutor con el que he tenido la suerte de trabajar, por toda la atención y ayuda que me ha prestado a lo largo de todo el proceso.

No puede faltar la mención a los grandes amigos que he conocido durante este tiempo y que me han sido de gran apoyo, con los que he compartido buenos momentos.

Por último y lo más importante, dar las gracias a mi familia. Todo lo conseguido ha sido posible gracias a ellos, que siempre han estado a mi lado y nunca ha dejado de creer en mí.

#### REFERENCIAS

- [1] "Introduction to Cisco IOS NetFlow." [Online]. Available: [http://www.cisco.com/c/en/us/products/collateral/ios-nx-os-software/ios-netflow/prod\\_white\\_paper0900aecd80406232.pdf](http://www.cisco.com/c/en/us/products/collateral/ios-nx-os-software/ios-netflow/prod_white_paper0900aecd80406232.pdf)
- [2] M. Roughan, A. Greenberg, C. Kalmanek, M. Rumsewicz, J. Yates, and Y. Zhang, "Experience in measuring backbone traffic variability: Models, metrics, measurements and meaning," *ACM SIGCOMM Internet Measurement Workshop*, pp. 91–92, 2002.
- [3] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg, "Fast accurate computation of large-scale IP traffic matrices from link loads," *ACM SIGMETRICS Performance Evaluation Review*, vol. 31, no. 1, p. 206, 2003.
- [4] M. Rahman, S. Saha, U. Chengan, and A. S. Alfa, "IP Traffic Matrix Estimation Methods : Comparisons and Improvements," vol. 00, no. c, pp. 90–96, 2006.
- [5] A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, and C. Diot, "Traffic Matrices : Balancing Measurements , Inference and Modeling," *ACM SIGMETRICS Performance Evaluation Review*, vol. 33, no. 1, pp. 362–373, 2005.
- [6] J. Camacho, A. Pérez-Villegas, P. García-Teodoro, and G. Maciá-Fernández, "PCA-based multivariate statistical network monitoring for anomaly detection," *Computers and Security*, vol. 59, pp. 118–137, 2016.

- [7] J. Camacho, "Technical Report : Networkmetrics . Multivariate Visual Analytics for Networking Data Multivariate Techniques based on Projection Subspaces," pp. 1–25, 2014.
- [8] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [9] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [10] "Internet2 Network." [Online]. Available: <http://www.internet2.edu>
- [11] "Internet2 Network Operation Center." [Online]. Available: <http://noc.net.internet2.edu>
- [12] E. Zhao and L. Tan, "A PCA based optimization approach for IP traffic matrix estimation," *Journal of Network and Computer Applications*, 2015.
- [13] H. Abdi, "Partial Least Squares (PLS) Regression," *Encyclopedia for research methods for the social sciences*, pp. 792–795, 2003.
- [14] S. Wold, M. Sjöström, and L. Eriksson, "Plsregression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, pp. 109–130, 2001.
- [15] P. Geladi and B. Kowalski, "Partial Least-Squares Regression: a Tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986.
- [16] L. E. Mujica, J. Rodellar, A. Fernández, and A. Güemes, "Q-statistic and T2-statistic PCA-based measures for damage assessment in structures," *Structural Health Monitoring*, vol. 10, no. 5, pp. 539–553, 2011.
- [17] J. A. Westerhuis, S. P. Gurden, and A. K. Smilde, "Generalized contribution plots in multivariate statistical process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 95–114, 2000.
- [18] J. E. Jackson and G. S. Mudholkar, "Control procedures for residuals associated with principal component analysis," *Technometrics*, vol. 21, no. 3, pp. 341–349, 1979.
- [19] J. A. Westerhuis, S. P. Gurden, and A. K. Smilde, "Standardized Q-statistic for improved sensitivity in the monitoring of residuals in MSPC," no. August 1999, pp. 335–349, 2000.
- [20] B. M. Wise, N. L. Ricker, D. F. Veltkamp, and B. R. Kowalski, "Pca monitoring theoretical basis," 1990.
- [21] K. Papagiannaki, N. Taft, and A. Lakhina, "A Distributed Approach to Measure IP Traffic Matrices," *Conference on Internet measurement*, pp. 161–174, 2004.

**Francisco José Cuberos Muñoz** nacido en Granada el 14/01/1992), graduado en Ingeniería de Tecnologías de Telecomunicación por la Universidad de Granada en 2015. Finalizó el Máster Universitario en Ingeniería de Telecomunicación en 2017. Actualmente cursa el Máster Propio en Ciberseguridad impartido en la Escuela Superior de Ingenierías Informática y Telecomunicación (ETSIT) de la Universidad de Granada.



**José Camacho Páez** El Dr. José Camacho es Profesor Titular de la Universidad de Granada. Pertenece a los grupos de investigación NESG (Network Engineering and Security Group) y UCyS (UGR Cyber Security Group) y es miembro del CITIC (Centro de Investigación en Tecnologías de la Información y la Comunicación). Es autor de más de 80 publicaciones, de las que 30 son artículos en revista internacional de impacto. Ha participado en más de 15 proyectos de investigación y contratos de transferencia, valorados en más de 1M€. Por su labor investigadora ha recibido varios premios nacionales e internacionales. Es profesor del Máster de Ciberseguridad de la Universidad de Granada. Ha sido o es supervisor de 4 Tesis Doctorales y más de 25 Trabajos Fin de Máster y Grado, y evaluador de 9 Tesis Doctorales. Es desarrollador de varias herramientas de software libre y socio fundador de dos startup.



# ESTIMACIÓN DE MATRICES DE TRÁFICO EN SERIES TEMPORALES CON MÍNIMOS CUADRADOS PARCIALES

Autor: Irene Herrera López, e-mail: irenehl@correo.ugr.es

Tutor: José Camacho Páez, e-mail: josecamacho@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—La cantidad de información que fluye entre todos los posibles pares origen y destino de una red durante un determinado periodo de tiempo se denomina matriz de tráfico (TM, Traffic Matrix). Son diversas las técnicas propuestas a lo largo de los años para su obtención, pero estas suponen distintos compromisos entre el *overhead* incurrido y la precisión de la estimación obtenida. Este Trabajo Fin de Máster parte de un trabajo colaborativo con Francisco José Cuberos, suponiendo la primera aplicación de la técnica de Mínimos Cuadrados Parciales (PLS, *Partial Least Squares*) al problema de la estimación de matrices de tráfico. Esta herramienta ha sido desarrollada ampliamente en campos como la econometría, la psicometría y la quimiometría pero, sin embargo, es una aplicación pionera en la estimación de matrices de tráfico. Tras obtener unos resultados prometedores en términos de rendimiento con PLS, en este Trabajo se ha implementado la combinación de PLS con modelos en series temporales para la optimización del compromiso entre el *overhead* y la precisión de la estimación.

**Palabras clave**—Matriz de Tráfico, estimación, Mínimos Cuadrados Parciales, *overhead*, Series Temporales

## I. INTRODUCCIÓN

**A** GROSSO modo, una matriz de tráfico es una representación abstracta del volumen de tráfico que fluye entre un conjunto de pares de nodos origen y destino. Cada elemento de la matriz simboliza la cantidad de tráfico entre un par origen y destino que, para abreviar, se denotará por OD (Origen-Destino). Se puede afirmar, por lo tanto, que la matriz de tráfico es un parámetro de entrada crucial para los distintos procesos de planificación de una red. El estudio de las matrices de tráfico, sumado al conocimiento de la topología y la información de *routing*, son de vital importancia para unas buenas prácticas de planificación y gestión de la red. Son multitud de tareas las que se benefician de este conocimiento, tales como la optimización, la detección de posibles anomalías y el diseño de protocolos.

Sin embargo, a pesar de su importancia, la obtención de la matriz de tráfico presenta una serie de problemas.

## II. ESTIMACIÓN DE MATRICES DE TRÁFICO

Se va a representar la TM como una serie temporal de matrices bidimensionales,  $\mathbf{X}$ . Cada elemento  $X_{n,n}$  denota la cantidad de tráfico entre un par OD. Dicha cantidad puede referirse a *bytes* o paquetes que han sido medidos durante

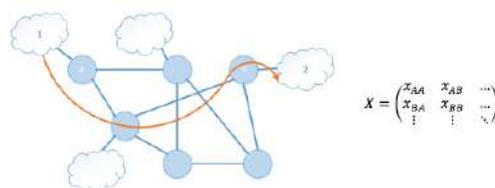


Fig. 1. Ejemplo matriz de tráfico.

un determinado intervalo de tiempo. En la figura 1 se ha representado, de manera abstracta, el concepto de TM.

Esta matriz cuenta con  $N^2$  pares OD para una red con  $N$  nodos. El método más común para estimar la TM viene dado por la siguiente ecuación lineal:

$$\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t) \quad (1)$$

siendo  $\mathbf{y}(t)$  el vector que describe el volumen de tráfico para cada uno de los enlaces,  $\mathbf{A}$  la matriz de *routing* y  $\mathbf{x}(t)$  la representación vectorizada de la matriz de tráfico  $\mathbf{X}$ . Se pueden definir las matrices  $\mathbf{X}$  e  $\mathbf{Y}$  como:

$$\left. \begin{array}{l} \mathbf{X} := \mathbf{x}(t) \\ \mathbf{Y} := \mathbf{y}(t) \end{array} \right\} \text{ para } t = 1, \dots, T \quad (2)$$

La matriz  $\mathbf{A}$  es de dimensión  $L \times N^2$ , siendo  $L$  el número de enlaces en la red y  $N^2$  el número de pares OD. En la mayoría de despliegues, la matriz  $\mathbf{A}$  es conocida y los valores de  $\mathbf{Y}$  son fácilmente observables mediante, por ejemplo, SNMP (*Simple Network Management Protocol*).

En cuanto a la matriz de tráfico,  $\mathbf{X}$ , puede ser medida directamente utilizando, por ejemplo, NetFlow. Sin embargo, la medición directa de la matriz completa consume un alto nivel de recursos, suponiendo un aumento del *overhead* para un escenario real. NetFlow consiste en un conjunto de servicios que permiten a los administradores de la red acceder a información de los flujos IP. Dicha información es recolectada por los elementos de la red (routers y switches) en los que se haya implementado NetFlow. Se puede definir un flujo como una secuencia unidireccional de paquetes que atraviesan un nodo de la red y comparten propiedades como las direcciones IP, los puertos origen y destino y el protocolo. Por otro lado,

SNMP es un conjunto de herramientas de gestión de red que permite definir y almacenar los contadores deseados y el intervalo de muestreo. A diferencia de SNMP, NetFlow es más pesado (más de un orden de magnitud en estado comprimido frente a SNMP sin compresión). Sin embargo, NetFlow incluye información sobre el origen y destino de los paquetes que SNMP no proporciona. Por esta razón, SNMP se puede utilizar junto a NetFlow para estimar la matriz de tráfico, mientras que NetFlow sirve para obtenerla directamente.

### III. MATERIALES Y MÉTODOS UTILIZADOS

#### A. Conjunto de datos utilizado

Para validar las propiedades de la metodología propuesta en este Trabajo, se ha utilizado un conjunto de datos reales correspondientes a la red Abilene (actualmente Internet2 [4]). La red Abilene contaba en 2004 con 12 routers, distribuidos por Estados Unidos. En la figura 2 se puede ver la topología de la red Abilene. Se han numerado tanto los *routers* como las demandas de tráfico de acuerdo a la fuente original de los datos. Para esta topología, se tiene un total de  $N^2 = 12^2 = 144$  pares OD.

#### B. Matriz de routing

Para el cálculo de la matriz de routing se va a asumir que dicha matriz  $\mathbf{A}$  es constante durante el intervalo de medidas. Sin embargo, en un escenario real en el cual se dispusiera de la matriz de *routing*, se deberían tener en cuenta posibles cambios en la misma debido a caídas de enlaces, establecimiento de nuevas rutas o el propio uso de OSPF (*Open Shortest Path First*). Para mapear el tráfico y formar la matriz  $\mathbf{A}^1$ , se va a recurrir a los pesos OSPF proporcionados junto al conjunto de datos.

#### C. Metodologías utilizadas

Para el caso de las medidas de NetFlow proporcionadas, el muestreo es periódico y los resultados son agregados por flujo cada cinco minutos. Esto supone un alto consumo de recursos de red, existiendo el riesgo de colapso de dichos recursos. Sin embargo, al ser medidas reales, presenta un mayor nivel de precisión. Se tomará la matriz de tráfico formada por las

<sup>1</sup>La matriz  $\mathbf{A}$  se utiliza para obtener  $\mathbf{Y}$  tanto en PLS como en PCA. Sin embargo,  $\mathbf{A}$  también se usa en PCA para la obtención del modelo de estimación mientras que en PLS no es necesaria.

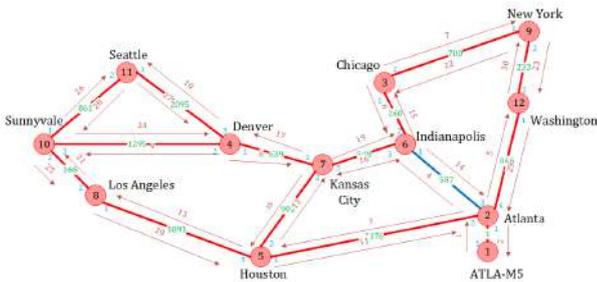


Fig. 2. Topología de la red Abilene.

mediciones de NetFlow como la TM real, a partir de la cual comenzarán nuestros experimentos.

La otra metodología seleccionada ha sido General Tomogravity [1] por ser uno de los métodos más completos. También se evaluará el rendimiento frente a la técnica de PCA [2] por ser una técnica predictiva ampliamente utilizada para la estimación de matrices de tráfico. La implementación de estas metodologías puede consultarse en la bibliografía seleccionada.

#### D. Métricas de rendimiento

Para cerrar este capítulo, se van a definir las métricas elegidas para evaluar el rendimiento del método propuesto. Una de las métricas más comúnmente utilizadas en la estimación de la TM es el error relativo medio. Se va a diferenciar entre error relativo medio espacial y error relativo medio temporal [5].

También se utilizará la función de distribución acumulativa (CDF, *Cumulative Distribution Function*) para cada tipo de error relativo medio: CSRE y CTRE para el SRE y el TRE, respectivamente. Así, se podrá observar la distribución de los errores.

1) *TRE (Temporal Relative Error)*: El error relativo medio temporal (TRE, *Temporal Relative Error*) para un tiempo  $t$  se define mediante la siguiente expresión [5]:

$$TRE(t) = \frac{\|\mathbf{X}^*(t) - \mathbf{X}(t)\|_2}{\|\mathbf{X}(t)\|_2} \quad (3)$$

donde  $\mathbf{X}^*(t)$  y  $\mathbf{X}(t)$  son los elementos estimados y reales, respectivamente, para un tiempo  $t$  dado. Representa el error de estimación de todos los flujos para un tiempo determinado.

2) *SRE Spatial Relative Error*: El error relativo medio espacial (SRE, *Spatial Relative Error*) del  $k$ -ésimo flujo OD se define mediante la siguiente expresión [5]:

$$SRE(k) = \frac{\|\mathbf{X}_T^*(k) - \mathbf{X}_T(k)\|_2}{\|\mathbf{X}_T(k)\|_2} \quad (4)$$

donde  $k = 1, \dots, N$ ,  $\mathbf{X}_T^*(k)$  y  $\mathbf{X}_T(k)$  son los elementos estimados y reales, respectivamente, del vector de tráfico del  $k$ -ésimo flujo OD durante un determinado periodo de tiempo con  $T$  muestras. Representa el error medio de estimación durante un periodo de tiempo para un flujo OD dado.

3) *Overhead*: El término *overhead* se refiere a la cantidad de información que necesita ser enviada a través de la red hasta la estación recolectora.

Con las alternativas propuestas, se pretende alcanzar un equilibrio entre *overhead* y precisión de la estimación obtenida. La estimación del *overhead* dependerá de la metodología utilizada. Teniendo esto en cuenta, se puede expresar el *overhead* para *General Tomogravity* y PLS/PCA, respectivamente, con las siguientes expresiones:

$$overhead_{GT} = T \cdot L_{GT} \quad (5)$$

$$overhead_{PLS} = T \cdot L_{PLS} + T' \cdot N^2 \quad (6)$$

donde  $T$  será el número de muestras temporales que conforman el intervalo de medición total, igual para ambas metodologías;  $L_{GT}$  y  $L_{PLS}$  será el número de enlaces por

los cuales se cursa el tráfico en cada metodología y  $N^2$  será el número de pares OD. Por último, la variables  $T'$  hace referencia al número de muestras seleccionadas tras aplicar el algoritmo de selección parcial con series temporales.

Así, para el caso de *General Tomogravity*, se deben tener en cuenta enlaces de tipo (*internal*, *inbound* y *outbound*). Sin embargo, para las metodología PLS [3] (al igual que PCA) se observan dos términos en las expresiones para la obtención del *overhead*. En el primero, sólo se tendrán en cuenta los enlaces de tipo *internal*, que son los que se han tenido en cuenta para la obtención de la matriz  $\mathbf{Y}$  a partir de la matriz de *routing*  $\mathbf{A}$ . El segundo término vendrá determinado por el algoritmo de series temporales empleado que definirá la reducción de muestras temporales para el total de pares OD que componen la red.

#### IV. MÍNIMOS CUADRADOS PARCIALES

Se parte del problema de la regresión lineal, el cual se define mediante la siguiente expresión:

$$\mathbf{X} = \mathbf{Y} \cdot \mathbf{B} + \mathbf{E} \quad (7)$$

donde  $\mathbf{X}^2$  es la matriz que se quiere estimar,  $\mathbf{Y}$  es la matriz de variables disponibles para realizar la estimación de  $\mathbf{X}$ ,  $\mathbf{B}$  es la matriz de coeficientes de regresión y  $\mathbf{E}$  es la matriz de residuos.

El método de mínimos cuadrados proporciona una solución al problema anterior que minimiza el error cuadrático. La solución de mínimos cuadrados para la ecuación 7 es:

$$\hat{\mathbf{B}}_{LS} = (\mathbf{Y}^T \cdot \mathbf{Y})^{-1} \cdot \mathbf{Y}^T \cdot \mathbf{X} \quad (8)$$

Puede darse el caso en el cual la matriz  $(\mathbf{Y}^T \cdot \mathbf{Y})$  sea singular, es decir, que su determinante sea cero y, por lo tanto, no se pueda invertir dicha matriz. En este caso, se dice que las variables de  $\mathbf{Y}$  son colineales. Por ello, no se puede estimar  $\mathbf{X}$  directamente a partir de  $\mathbf{Y}$ . Para solventar los problemas de colinealidad, es necesario seleccionar un número reducido de variables (obtenidas como la combinación lineal de las variables originales), denominadas variables latentes (LVs, *Latent Variables*).

Así, la alternativa propuesta por el método PLS es predecir  $\mathbf{X}$  a partir de un número determinado de LVs de  $\mathbf{Y}$ . Es decir, estimar  $\mathbf{X}$  mediante el subespacio de  $\mathbf{Y}$  que maximiza su covarianza con  $\mathbf{X}$ . El problema de la regresión parcial lineal entre las matrices normalizadas  $\mathbf{X}$  e  $\mathbf{Y}$  se puede definir de la siguiente manera [8]:

$$\mathbf{Y} = \mathbf{T}_A \cdot \mathbf{P}_A^T + \mathbf{E}_A \quad (9)$$

$$\mathbf{X} = \mathbf{T}_A \cdot \mathbf{Q}_A^T + \mathbf{F}_A \quad (10)$$

donde  $\mathbf{T}_A$  es la matriz  $N \times A$  de *scores*, la cual contiene las proyecciones de  $\mathbf{Y}$  al subespacio de dimensión  $A$ , siendo  $A$  el número de LVs seleccionado. Por otro lado,  $\mathbf{P}_A$  y  $\mathbf{Q}_A$  son las matrices de *loadings*, de dimensiones  $M \times A$  y  $K \times A$  respectivamente. Finalmente, las matrices  $\mathbf{E}_A$  y  $\mathbf{F}_A$

<sup>2</sup>En la literatura de PLS, típicamente, la notación de  $\mathbf{X}$  e  $\mathbf{Y}$  está invertida

corresponden a las matrices de residuos. Las ecuaciones 9 y 10 se pueden expresar como:

$$\mathbf{X} = \mathbf{Y} \cdot \mathbf{R}_A \cdot \mathbf{Q}_A^T + \mathbf{F} \quad (11)$$

con

$$\mathbf{R}_A = \mathbf{W}_A \cdot (\mathbf{P}_A^T \cdot \mathbf{W}_A)^{-1} \quad (12)$$

siendo  $\mathbf{W}_A$  la matriz  $M \times A$  de pesos.

Finalmente, se tiene la expresión para la matriz  $\mathbf{B}$ :

$$\mathbf{B}_{PLS} = \mathbf{R}_A \cdot \mathbf{Q}_A^T \quad (13)$$

#### A. Selección del número de Variables Latentes con Validación Cruzada

El conjunto de LVs se puede considerar como combinaciones lineales de los elementos de  $\mathbf{Y}$ . Las LVs serán los nuevos predictores de  $\mathbf{X}$ . Sin embargo, no todas las variables latentes serán necesarias para determinar el modelo de estimación final. Por ello, es necesario analizar el poder predictivo de cada LV para detectar aquellas que no sean significativas.

La validación cruzada es una buena opción para comprobar el poder predictivo mencionado. La métrica denominada PRESS (*Predictive Residual Sum of Squares*) nos informa sobre la capacidad predictiva del modelo. Uno de los métodos más sencillos de validación cruzada es la técnica *rkf* (*Row-wise k-fold*) [6]. En la figura 3 se muestra un ejemplo de la representación de la curva PRESS en función del número de LVs, representado en el eje de abscisas. Se elegirá como número óptimo de LVs aquel punto de la curva a partir del cual el valor de PRESS comienza a aumentar de nuevo tras alcanzar un mínimo. En caso de no alcanzar un mínimo, se toma el número de LVs para el que se estabiliza el valor de PRESS, es decir, a partir del cual no se obtiene una mejora significativa.

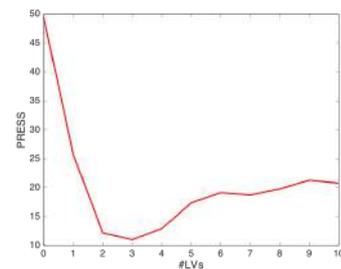


Fig. 3. Ejemplo del resultado de la validación cruzada *rkf*.

Una vez aplicada la metodología de validación cruzada, el número de LVs elegido,  $A$ , será aquel que proporcione un valor mínimo de PRESS. En el ejemplo mostrado en la figura 3 el valor elegido será  $A = 3$ . Esta selección se realiza con el conjunto de datos de calibración y se aplica a la estimación de la matriz  $\mathbf{B}$  con el conjunto de datos de test.

V. EXPERIMENTACIÓN

A. Tratamiento de los datos

Con el objetivo de validar la predicción con PLS, se dividen los datos en parte de calibración y test. Se va a dividir el conjunto de datos total de 19 semanas en dos partes. La primera parte (abarca las primeras diez semanas) se utilizará para elegir el algoritmo de selección parcial más adecuado. Por otro lado, se utilizarán las últimas 10 semanas (se utiliza la semana diez tanto para calibración como para test) para evaluar el rendimiento del algoritmo de selección. A su vez, dentro de cada uno de los bloques de diez semanas, se utilizará (para cada iteración) un subconjunto de calibración y otro de test. Con este otro conjunto de calibración se obtendrá el valor del número de LVs que se va a utilizar en la predicción, sobre los datos de test. En la figura 4 se ha esquematizado lo descrito anteriormente. Esto está motivado por el carácter cicloestacionario semanal del tráfico. En sucesivas iteraciones, se irá desplazando la ventana de calibración de día en día hasta cubrir la totalidad de los datos disponibles.

La matriz **X** estará formada por las mediciones obtenidas a partir de NetFlow, a la cuales se añadirá un cierto nivel de ruido, ya que esto aporta un mayor realismo al escenario de pruebas.

En cuanto a la matriz **Y**, se puede calcular directamente a partir de la matriz **X** (sin ruido añadido) y la matriz de *routing* **A**, según la siguiente expresión:

$$Y = XA^T \tag{14}$$

Las matrices **X** e **Y** correspondientes a cada iteración serán la entrada para el cálculo de la matriz **B**.

B. Implementación PLS

En esta primera fase de la experimentación se va a evaluar la viabilidad de usar PLS para estimar la matriz de tráfico a partir de la matriz **Y**, lo que supone un aumento del *overhead*. Esta metodología se va a denotar como PLS *offline*. Se va a valorar una aplicación alternativa de PLS, la cual se denominará PLS *online*, en la cual se pretende minimizar el *overhead* asociado mediante la reducción del número de mediciones de NetFlow.

Para la implementación del algoritmo PLS se ha utilizado la *toolbox* MEDA (*Multivariate Exploratory Data Analysis*) [7] para Matlab. Esta herramienta permite encontrar las relaciones existentes entre las características de los datos [8].

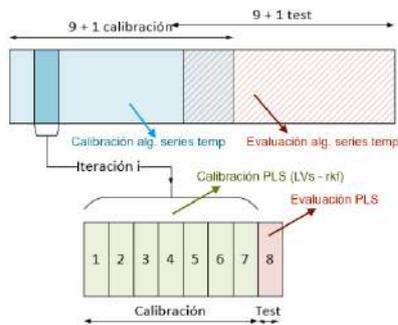


Fig. 4. División del conjunto de datos en grupos de calibración y test.

1) *Algoritmo 1: Serie temporal con selección parcial de los datos*: La primera propuesta consiste en seleccionar la frecuencia adecuada para tomar las medidas. Para ello, se define un parámetro denominado  $C_t$ , que indicará cada cuanto tiempo se realizarán medidas.

Se ha realizado un estudio para encontrar el valor de  $C_t$  óptimo el cual nos permita alcanzar un nivel de *overhead* similar al que se obtiene con *General Tomogravity*. Así, se tratará de alcanzar un equilibrio entre *overhead* y precisión de la estimación obtenida con PLS para distintos escenarios. Finalmente, se ha seleccionado el valor  $C_t = 6$  como aquel valor óptimo que proporciona un *overhead* similar a la referencia de *General Tomogravity*.

En la figura 5 se ha representado la evolución de los distintos parámetros que se han presentado en la tabla ?? en función del valor de  $C_t$ , representado en el eje de abscisas para las tres sub-gráficas. En la sub-figura 5 (a), el eje de ordenadas representa el valor SRE medio obtenido para las estimaciones. Se han establecido dos umbrales: el SRE medio para *General Tomogravity* (color rojo) y el SRE medio para PLS *offline* (color verde). Estos valores son constantes ya que no dependen del valor  $C_t$ . Sin embargo, se puede observar cómo, al aumentar el valor de  $C_t$ , se incrementa al SRE, es decir, al disminuir el número de muestras seleccionadas, la estimación obtenida es menos precisa. La interpretación de la sub-figura 5 (b) es semejante a la anterior pero, en este caso, se representa el TRE medio en el eje de ordenadas. En la última sub-figura 5 (c), se muestra el *overhead* en el eje de ordenadas. Como se puede observar, el valor del *overhead* disminuye al incrementar el valor de  $C_t$ , es decir, al disminuir el número de muestras recolectadas. Se ha marcado en rojo el punto en el cual el *overhead* del algoritmo propuesto alcanza el *overhead* de *General Tomogravity* para  $C_t = 6$ . Se aprecia gráficamente la disminución considerable respecto al *overhead* de la metodología PLS *offline*. También se observa una reducción en el TRE y SRE respecto a *General Tomogravity*.

De esta manera, se está reduciendo considerablemente el *overhead* con un pequeño sacrificio en términos del error. A continuación, una vez elegido el valor óptimo de  $C_t$ , se va a aplicar sobre la segunda mitad del conjunto de datos para comprobar el rendimiento de este primer algoritmo propuesto. Los resultados numéricos se pueden observar en la tabla I, en

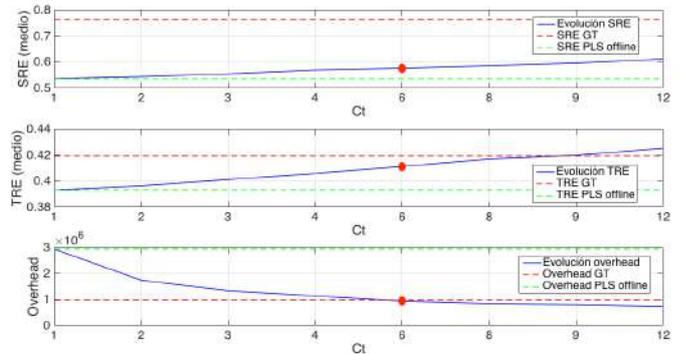


Fig. 5. Representación de la evolución de TRE, SRE y del *overhead* para los distintos casos de  $C_t$  en función de *General Tomogravity* y PLS *offline*

la cual se han incorporado también los valores obtenidos para la metodología PCA (bajo las mismas condiciones). Como se ve, se obtienen mayores valores del error para  $C_t = 6$  que con la versión *offline* (tanto en PLS como PCA), y a que al no tener en cuenta todas las muestras la estimación no es tan precisa. Sin embargo, la metodología propuesta para PLS con series temporales y  $C_t = 6$  presenta un mejor rendimiento que *General Tomogravity* y que PCA para un nivel de *overhead* similar.

De manera gráfica, se han representado los valores del TRE y SRE en las figuras 6 y 7, respectivamente.

En la figura 6 (a) se ha representado el valor de TRE en función del tiempo. Se puede observar, por ejemplo, que los valores del TRE para PLS con  $C_t = 6$  (color morado) son ligeramente inferiores a los del TRE para PCA con  $C_t = 6$  (color verde). En la figura 6 (b) se observa en el eje de ordenadas la CDF en función del TRE representado en el eje de las abscisas. En esta gráfica se observa cómo la curva que queda por encima del resto es la correspondiente a la metodología PLS *offline*, seguida por la curva de PLS con  $C_t = 6$ , superando el rendimiento en función del TRE de las metodologías *General Tomogravity* y las dos versiones de PCA, *offline* y  $C_t = 6$ .

En la figura 7 se puede aplicar el mismo razonamiento que en la figura anterior, en este caso para el valor de SRE. En la primera sub-figura se han resaltado los pares OD ejemplo en los que se consigue reducir considerablemente el valor del SRE respecto a *General Tomogravity offline* y  $C_t = 6$ . Las curvas que representan la CDF para el SRE no son tan claras, pero puede apreciarse cómo la curva correspondiente a PLS con  $C_t = 6$  permanece por encima de la curva de la metodología *General Tomogravity* para la mayor parte de los valores del SRE.

Tabla I  
RESULTADOS FINALES DE LA EVALUACIÓN DEL ALGORITMO 1.

	GT	PLS <i>offline</i>	PLS $C_t = 6$	PCA <i>offline</i>	PCA $C_t = 6$
TRE	0.4474	0.3971	0.4177	0.4780	0.4784
SRE	0.8456	0.5146	0.5655	0.5697	0.5816
Overhead (B)	979776	2939328	943488	2939328	943488

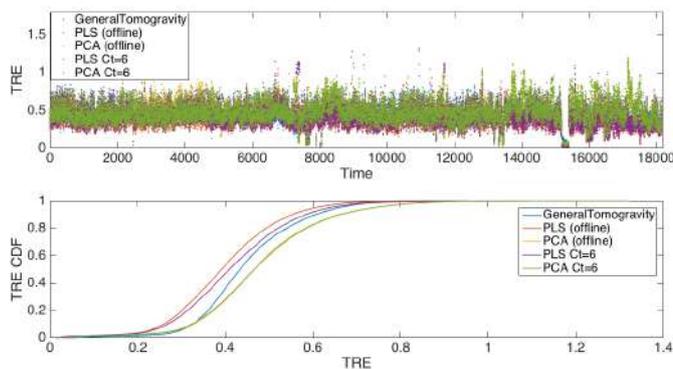


Fig. 6. Representación del TRE para el algoritmo 1 de medición parcial con  $C_t = 6$ .

2) *Algoritmo 2: Serie temporal con realimentación de la estimación y selección parcial de los datos:* En esta segunda alternativa se va a proponer la realimentación del algoritmo con las estimaciones calculadas en el instante anterior. En la figura 8 se ha esquematizado la metodología utilizada.

En este caso, llegará una iteración para la cual la ventana de calibración esté compuesta en su totalidad por estimaciones previas. Esto puede provocar que los errores de la estimación final aumenten hasta que el rendimiento no se considere aceptable, debido a la continua reintroducción de los errores propios de las estimaciones. Por lo tanto, se propone añadir muestras de NetFlow a las estimaciones que se reincorporan a la ventana de calibración, con el objetivo de encontrar un equilibrio entre reducción del *overhead* y la precisión de la estimación final. Es decir, la realimentación de la ventana de calibración estará conformada por una parte de la estimación y por otra parte de NetFlow real. Las muestras de NetFlow que van a sustituir a las muestras de la estimación deben elegirse de manera que tengan correspondencia temporal. Este parámetro se va a denominar  $C_t$ , que indicará la cantidad de muestras que se van a tomar en cada medición. También se va a definir un parámetro adicional,  $\Delta_t$ , el cual definirá cada cuánto tiempo se realizan las medidas.

Se ha implementado este segundo algoritmo para las distintas combinaciones de  $C_t = 1, 2, 4, 8$  y  $\Delta_t = 288, 144, 72, 36$ . Se ha elegido  $C_t = 4$  y  $\Delta_t = 72$  ya que es la que menor TRE medio presenta para un valor de *overhead* semejante al de *General Tomogravity*. Aunque el valor de SRE medio no es el menor para esta combinación, éste no se ha tenido en cuenta para la elección, ya que el SRE puede presentar

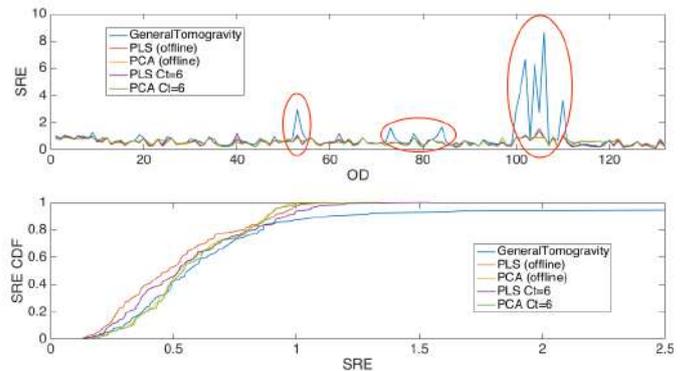


Fig. 7. Representación del SRE para el algoritmo 1 de medición parcial con  $C_t = 6$ .

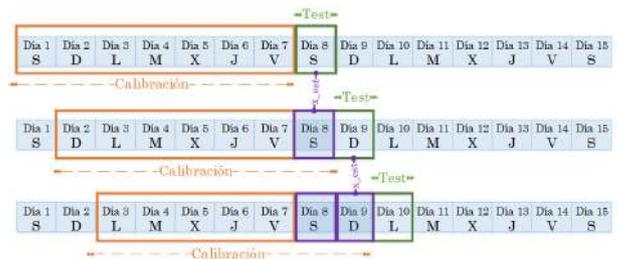


Fig. 8. Representación del algoritmo 2 de serie temporal con reintroducción de la estimación y selección parcial de los datos.

outliers o valores atípicos debido a los flujos pequeños, que pueden afectar al conjunto.

Para finalizar la experimentación de este segundo algoritmo propuesto, se va a estimar la TM para la segunda mitad de los datos con los parámetros  $C_t = 4$  y  $\Delta_t = 72$  tanto con el modelo PLS como con PCA. En la tabla II se pueden ver los resultados del rendimiento. En dicha tabla se representan los valores medios de los errores obtenidos para todas las metodologías objeto de comparación de este estudio.

Para finalizar este capítulo y, a modo de resumen, se muestran en la tabla III los resultados finales de los errores medios para el TRE y el SRE, así como el valor de *overhead* para las metodologías de *General Tomogravity*, PLS y PCA. Para el caso de PLS y PCA se recogen los datos para sus tres modalidades presentadas: *offline* en la cual se recogen todas las muestras temporales, Alg 1 corresponde al primer algoritmo con  $C_t = 6$  y Alg 2 se refiere al segundo algoritmo con realimentación de la estimación con  $C_t = 4$  y  $\Delta_t = 72$ .

Los errores obtenidos para el algoritmo 2 son algo superiores a los resultados del algoritmo 1:  $TRE_1 = 0.4177$  y  $SRE_1 = 0.5655$  frente a  $TRE_2 = 0.4373$  y  $SRE_2 = 0.6251$ . Esto es debido a la realimentación del algoritmo 2, ya que los errores cometidos en la estimación para cada iteración son reintroducidos en el modelo. Aún así, el rendimiento medio de la metodología PLS *online* es mejor respecto a los errores de estimación obtenidos con *General Tomogravity*. Además, y como cierre a este capítulo, la significativa reducción del *overhead* obtenido con PLS *online* frente a PLS *offline* proporciona una ventaja real para la implementación de esta metodología en escenarios reales.

## VI. CONCLUSIONES

Para concluir este Trabajo, cabe destacar que la solución obtenida no sólo ha cumplido los requisitos iniciales, sino que, además, se han obtenido mejoras respecto a otras técnicas del Estado del Arte de la estimación de matrices de tráfico. En resumen, se ha optimizado el compromiso entre el *overhead*

y la precisión de la estimación y se han obtenido mejores estimaciones, en términos de precisión, que otras técnicas al mismo nivel de *overhead*. Por otro lado, se debe tener en cuenta que la optimización de dicho compromiso resulta indispensable para la aplicación de la técnica de PLS en un escenario real de estimación de matrices de tráfico. Sumado a todo lo anterior, hay que tener en cuenta la aplicación pionera de la técnica PLS en el campo de la estimación de matrices de tráfico.

## AGRADECIMIENTOS

Agrader a Pepe sus consejos, esfuerzo y la pasión por la investigación. Dar las gracias también a Franci y a mi familia por todo el apoyo.

## REFERENCIAS

- [1] Y. Zhang, M. Roughan, N. Duffield, A. Greenberg, "Fast Accurate Computation of Large-Scale IP Traffic Matrices from Link Loads".
- [2] J. Camacho, A. Pérez-Villegas, P. García-Teodoro, G. Maciá-Fernández, "PCA-based multivariate statistical network monitoring for anomaly detection".
- [3] S. Wold, M. Sjöström, L. Eriksson, "PLS-regression: a basic tool of chemometrics".
- [4] Internet2 Network, <http://www.internet2.edu>.
- [5] E. Zhao, L. Tan, A PCA based optimization approach for IP traffic matrix estimation, Journal of Network and Computer Applications.
- [6] J. Camacho, A. Ferrer, "Cross-validation in PCA models with the element-wise k-fold (ekf) algorithm: Practical Aspects".
- [7] J. Camacho, A. Pérez, R. Rodríguez, E. Jiménez-Mañas, "Multivariate Exploratory Data Analysis (MEDA) Toolbox". Disponible: <https://github.com/josecamachop/MEDA-Toolbox>
- [8] J. Camacho, "Technical Report : Networkmetrics . Multivariate Visual Analytics for Networking Data Multivariate Techniques based on Projection Subspaces" (2014) 1-25.

**Irene Herrera López** Graduada en Ingeniería de Tecnologías de Telecomunicación con mención en Telemática y Máster en Ingeniería de Telecomunicaciones. Actualmente, estudiante del Máster Propio en Ciberseguridad de la Universidad de Granada.



**José Camacho Páez** El Dr. José Camacho es Profesor Titular de la Universidad de Granada. Pertenece a los grupos de investigación NESG (Network Engineering and Security Group) y UCyS (UGR Cyber Security Group) y es miembro del CITIC (Centro de Investigación en Tecnologías de la Información y la Comunicación). Es autor de más de 80 publicaciones, de las que 30 son artículos en revista internacional de impacto. Ha participado en más de 15 proyectos de investigación y contratos de transferencia, valorados en más de 1M€. Por su labor investigadora ha recibido varios premios nacionales e internacionales. Es profesor del Máster de Ciberseguridad de la Universidad de Granada. Ha sido o es supervisor de 4 Tesis Doctorales y más de 25 Trabajos Fin de Máster y Grado, y evaluador de 9 Tesis Doctorales. Es desarrollador de varias herramientas de software libre y socio fundador de dos start-up.



Tabla II  
VALORES MEDIOS DE LOS ERRORES RELATIVOS CORRESPONDIENTE A LAS 10 ÚLTIMAS SEMANAS DEL CONJUNTO DE DATOS PARA EL ALGORITMO 2.

	GT	PLS <i>offline</i>	PLS $C_t = 4$ $\Delta_t = 72$	PCA <i>offline</i>	PCA $C_t = 4$ $\Delta_t = 72$
TRE	0.4474	0.3971	0.4373	0.4780	0.4917
SRE	0.8456	0.5146	0.6251	0.5697	0.6848
Overhead (B)	979776	2939328	941376	2939328	941376

Tabla III  
RESULTADOS FINALES DEL RENDIMIENTO DE LAS METODOLOGÍAS PROPUESTAS.

	TRE	SRE	Overhead (B)
<i>General Tomogravity</i>	0.4474	0.8456	979776
PLS	<i>Offline</i>	0.3971	2939328
	Alg 1	0.4177	943488
	Alg 2	0.4373	941376
PCA	<i>Offline</i>	0.4780	2939328
	Alg 1	0.4784	943488
	Alg 2	0.4917	941376

# Aplicación de Realidad Virtual con control cefálico: SkyMet VR

Autor: Gabriel Ávila Muñoz, e-mail: gabri95@correo.ugr.es

Tutor: Miguel Ángel López Gordo; e-mail: malg@ugr.es

Titulación: Máster en Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones

Universidad de Granada

**Resumen**— El crecimiento de la realidad virtual ha conseguido que esta tecnología sea más accesible y con ello se han generado multitud de aplicaciones. Una de ellas es su uso para tratamientos fisioterapéuticos que hacen más sencillo realizar las tareas y que pueden aportar datos más precisos de la evolución de los pacientes a los profesionales que los cuidan. Con datos más exactos sobre el impacto de la terapia en el paciente, los profesionales pueden gestionar mejoras en las terapias a la vez que los usuarios realizan con mayor ánimo sus tratamientos al estar inmersos en otra tarea más allá de la terapia. Su aplicación a personas con discapacidades neuronales cobra bastante importancia respecto a otro tipo de rehabilitación más leve donde también tiene utilidad esta tecnología.

Este proyecto consiste en la creación de un videojuego basado en realidad virtual para su uso en rehabilitación de personas con problemas troncoencefálicos, es decir, es un videojuego controlado por el movimiento de la cabeza. Responde a la necesidad de facilitar la vida a estas personas, especialmente cuando son niños con problemas neurológicos importantes tipo parálisis cerebral infantil los que lo sufren, ya que el uso de este proyecto busca el doble propósito de entretenerlos mientras realizan sus terapias a la vez que motivarlos para que sigan mejorando.

**Palabras clave**—Realidad Virtual, Rehabilitación, Motivación, Videojuego, Oculus, EnPathia

## I. INTRODUCCIÓN

CADA año se hace más evidente que tanto la realidad aumentada como la realidad virtual han llegado para quedarse. Esto se debe a que son tecnologías llamativas que tienen potencial para aportar muchas soluciones en una gran variedad de ámbitos y que cada vez son más asequibles de manera que es más sencillo que lleguen al ámbito doméstico.

Sin embargo, Realidad Aumentada y Realidad Virtual no son la misma tecnología aunque tengan sus semejanzas. El fundamento de ambas consiste en aportar información adicional a nuestros sentidos con la ayuda de algún dispositivo. La principal diferencia es que la realidad aumentada aporta dicha información superpuesta al mundo real, de manera que no perdemos esa referencia. Por el contrario, la realidad virtual nos sumerge en un mundo totalmente virtual donde nada de lo que percibimos realmente se encuentra donde lo percibimos.

El otro punto importante del proyecto es su aplicación al

mundo de la rehabilitación, más concretamente a la parte de enfermedades neurológicas. Existen más de 600 enfermedades neurológicas las cuales se han clasificado en distintos tipos. De todos ellos, si nos centramos en los que tienen relevancia en la movilidad cervical, nos encontramos con los siguientes tipos [1]:

- Condiciones traumáticas que implican a los sistemas óseo y ligamentoso de la columna cervical.
- Afecciones degenerativas y reumáticas.
- Condiciones neurológicas y congénitas.
- Limitaciones en la movilidad cervical relacionadas con el dolor debido a condiciones relacionadas con la postura, el trabajo o el estrés.

Las patologías más comunes son los traumatismos troncoencefálicos (del primer tipo) y la parálisis cerebral infantil (perteneciente al tercer tipo).

## II. OBJETIVOS

El objetivo de este proyecto es el desarrollo de un videojuego basado en realidad virtual que pueda ser utilizado en terapias de rehabilitación por personas con problemas de movilidad troncoencefálica de manera que jugando ejecuten las tareas y movimientos correspondientes. Es por tanto imperativo que el movimiento del personaje del juego se realice mediante movimiento cefálico del paciente y que el propio juego tenga una capacidad adaptativa a cada uno de ellos.

Todo ello debe de ser guiado por el fisioterapeuta correspondiente que adicionalmente va a recibir una respuesta por parte del juego que le permita evaluar el rendimiento del paciente y de esta manera pueda tomar decisiones que le permitan mejorar la terapia.

## III. TRABAJOS PREVIOS

Ante el éxito que iban sumando tanto la realidad aumentada como la realidad virtual, muchos profesionales se propusieron aplicarla al ámbito de la rehabilitación. Hay muchos trabajos al respecto y yo los he agrupado en tres tipos de soluciones:

- Proyectos con sensores y monitor. Es el grupo más numeroso dado que es el más sencillo de llevar a cabo.

Consiste en la utilización de sensores que tomen datos de la parte que se esté tratando del paciente y muestren alguna interacción con algo que ocurre en un monitor. En un trabajo destacado de este grupo utilizan un sensor Leap Motion que capta movimiento de los dedos de las manos y las propias manos [2].

- Proyectos con realidad aumentada. Es un nivel más en la inmersión del paciente en su tarea. El paciente interactúa con elementos que no se encuentran físicamente en el espacio donde se encuentra. La atención que genera esto hace que el paciente se centre más en los elementos irreales y realice los movimientos deseados de manera más instintiva. Entre los dispositivos más usados por trabajos de este grupo se encuentra el sensor Kinect de Xbox. En uno de esos trabajos se simulaban topes que el paciente debía pisar [3].

- Proyectos con realidad virtual. Es el nivel máximo de abstracción ya que el paciente no percibe el mundo real, por lo que es consciente por medio de sus propios sentidos de los movimientos que realiza. Esto ayuda a desarrollar la propiocepción, es decir, la consciencia de los movimientos y posiciones de los elementos de nuestro cuerpo sin usar los sentidos. Éste es el grupo con menor cantidad de trabajos porque es una tecnología más compleja de desarrollar y generalmente requiere de mayor presupuesto. Para desarrollarlos hace falta un visor de realidad virtual. Uno de estos trabajos desarrolla un ambiente virtual en donde los movimientos que realiza el paciente buscan recuperar la plasticidad cerebral [4].

Se han llevado a cabo diversos estudios [5-8] con distintos tipos de pacientes que tratan de establecer hasta qué punto puede ser beneficioso la aplicación de estas tecnologías a la rehabilitación. La conclusión es que promete serlo aunque aún no existen suficientes evidencias que lo respalden. Hacen falta más prototipos y más estudios.

#### IV. REQUISITOS

Este proyecto ha tenido una orientación empresarial y se ha contado con la participación del fisioterapeuta D. José Amador de la clínica Aúrea de Granada como cliente, de manera que ha colaborado a la hora de especificar una serie de requisitos, probar los prototipos y generar una valoración. De acuerdo a ello, los requisitos de un prototipo que cumpla con el objetivo marcado son:

- El movimiento del personaje debe de ser controlado por medio de control cefálico del paciente.
- El movimiento que debe de ejercitar el paciente depende de la patología asociada.
- El videojuego tiene que tener una capacidad adaptativa puesto que no todos los pacientes tienen el mismo grado de dolencia.
- El videojuego no debe contar con una ambientación estresante ya que los distintos pacientes capacidad sensorial diferente, como por ejemplo aquellos con parálisis cerebral infantil.
- La tecnología utilizada debe de ser confortable para el paciente.
- El videojuego debe aportar datos acerca del rendimiento del paciente.

#### V. DISEÑO E IMPLEMENTACIÓN

De acuerdo a los objetivos y requisitos del proyecto, se ha diseñado un sistema en el que el videojuego es la parte fundamental. Se ha desarrollado para dos plataformas distintas pensando en la confortabilidad del paciente, de manera que si no se siente cómodo con la realidad virtual, que es la plataforma principal, pueda utilizar otra alternativa perteneciente al grupo de proyectos con sensor más monitor. Adicionalmente se han incluido dos modos de juego pensando en las distintas patologías y necesidades que puedan presentar los pacientes.

##### A. Componentes

Para el desarrollo de la plataforma con sensor se ha utilizado el sensor enPathia, desarrollado por la empresa Eneso, que es un sensor que se coloca en una cinta alrededor de la cabeza y lee los movimientos de la misma. En el caso de la realidad virtual se han utilizado las gafas de realidad virtual Oculus Go que tienen procesador propio y sensores que permiten detectar los movimientos de la cabeza. En ambos casos el videojuego ha sido desarrollado mediante el motor de videojuegos Unity.

##### B. Sistema

El sistema está formado por tres entidades bien diferenciadas:

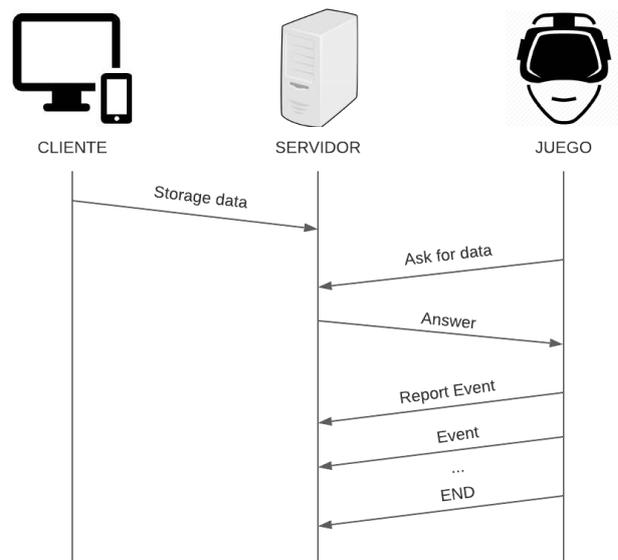


Fig. 1. Mensajes del protocolo SMCP entre las distintas entidades del sistema

- Cliente. Entidad manejada por el fisioterapeuta. Le permite adaptar el videojuego al paciente estableciendo una serie de parámetros que almacena en un archivo que posteriormente carga en el servidor.
- Servidor. Almacena los archivos de los pacientes y carga el seleccionado por el cliente. Cuando recibe una petición de la plataforma de juego le sirve el contenido de este archivo. También se encarga de recoger los eventos que ocurren en el juego que la plataforma le envía y los guarda en un archivo.
- Juego. Consiste en la plataforma concreta que ejecuta el juego. Recibe los parámetros designados por el cliente por

medio del servidor y durante la ejecución del videojuego envía al servidor los eventos que van ocurriendo.

Para la comunicación entre las distintas entidades se ha desarrollado un protocolo propio que se ha denominado SkyMet Communication Protocol. En la figura 1 se muestra un diagrama de intercambio de mensajes del protocolo. El sistema consta de tres etapas:

La primera consiste en el establecimiento de los datos por parte del fisioterapeuta en un archivo de paciente. Dicho archivo se almacena en el servidor y antes de comenzar un nuevo juego, el fisioterapeuta desde el cliente le indica al servidor que ficha de paciente ha de cargar. Esto lo realiza mediante el mensaje del protocolo 'Storage data'.

La segunda etapa viene determinada por la plataforma que ejecuta el juego cuando le envía el mensaje 'Ask for data' al servidor preguntándole si hay algún fichero de paciente que deba cargar antes de iniciar el juego (realmente esto se hace cuando el paciente está jugando el tutorial antes de iniciar el juego propiamente dicho). A este mensaje el servidor genera una respuesta con los parámetros obtenidos de la ficha de paciente seleccionada que debe cargar el juego. Si la respuesta llega sin parámetros, el videojuego entiende que no hay un paciente específico para jugar y carga parámetros por defecto.



Fig. 2. Presentación del juego SkyMet VR

Finalmente, la tercera etapa consiste en el envío de mensajes de evento con los sucesos que experimenta el paciente mientras juega. Esta etapa está enmarcada entre los mensajes 'Report Event', en el que el juego le indica al servidor que va a comenzar a enviar eventos, y 'END', en el que el juego le indica al servidor que ha terminado la partida y que por tanto puede guardar y cerrar el archivo con los eventos que ha ido almacenando.

Todos los mensajes del protocolo SMCP están formados por cadenas de string codificadas en UTF-8, de manera que para símbolos pertenecientes al conjunto US-ASCII (alfabeto inglés, números y caracteres más comunes como espacio, coma o punto) cada uno se codifica con un único byte. Pensando en un futuro para el protocolo, se le ha definido a cada mensaje una longitud de hasta 1024 bytes, aunque actualmente en ningún caso llega a esa capacidad.

### C. El juego

El diseño del videojuego se ha basado en el género calificado

como 'endless runner' donde el jugador conduce un personaje en carrera infinita en la que van apareciendo obstáculos que esquivar y ventajas que coger. En este tipo de juegos se permite un movimiento horizontal del personaje dentro de los límites del camino que recorren y movimientos horizontales tipo salto o agachar el personaje. Esos movimientos simples convierten este género en ideal para el objetivo de este proyecto, ya que se pueden asociar fácilmente al ejercicio troncoencefálico. Sin embargo, es interesante el no estar condicionado a correr sobre un camino que limite los movimientos verticales, sino que se requiere de más libertad para una mayor ejercitación, y en consonancia, y basándose en juegos clásicos, se ha diseñado como personaje una nave y como obstáculos asteroides que se han de esquivar.

De acuerdo a todo ello, la mecánica del juego es simple: el jugador conduce la nave espacial mediante movimientos de cabeza (la extensión y flexión genera los movimientos verticales y la inclinación de la cabeza genera a su vez la inclinación de la nave) y los asteroides se van acercando más veloces y pueden ser más grandes a cada nivel que se supera. La puntuación irá aumentando por cada asteroide que se esquivé, teniendo en cuenta el tamaño del meteorito que se esquivó. El jugador posee un máximo de 5 vidas que irá perdiendo cuando colisione con un meteorito. Como ventajas se han incluido corazones que al cogerlos recuperan una vida y estrellas que le permiten al jugador doblar la puntuación. Sin embargo, el verdadero atractivo del videojuego se encuentra en un superbonus, es decir, el disparo de misiles que localizan los asteroides que haya en el mapa y los destruye generando mayor puntuación que al esquivarlos. El objetivo del juego es que el paciente dispare tanto como sea posible estos misiles. Adicionalmente, el disparo de los misiles está condicionado al movimiento que el fisioterapeuta quiera potenciar para un determinado paciente. Por esta razón, existen dos modos de juego, y cada uno de ellos genera el disparo de misiles de distinta manera:



Fig. 3. Interfaz gráfica modo Estabilidad

El modo Estabilidad está pensado para persona que tienen problemas con la estabilidad de la cabeza, es decir, suelen tener la cabeza flexionada y les cuesta mantenerla erguida fijando la vista en el frente. Es por ello que en este modo de juego los misiles se dispararán cuando el jugador mantenga la cabeza durante un número determinado de segundos mirando al frente. El número de segundos depende del nivel de manera que cuanto más avanzado más segundos habrá de mantener la cabeza erguida. En este modo de juego todas las ventajas aparecen en la mitad superior de la pantalla para forzar un movimiento de

extensión cefálica por parte del paciente.



Fig. 4. Interfaz gráfica modo Progresión Angular

El modo Progresión Angular está pensado para gente que no tiene tanta capacidad articular y por lo tanto tiene que desarrollar más movimiento angular. Los misiles en este caso se dispararán al coger cierto número de estrellas que van apareciendo en el mapa alternativamente en la mitad superior e inferior del mismo. El conteo de estrellas a coger se realiza mediante cuenta atrás que aparece indicada en la interfaz de usuario, de manera que cuando se llegue a cero se dispararán los misiles y se comenzará la cuenta atrás de nuevo, aunque una unidad superior a la última cuenta atrás.

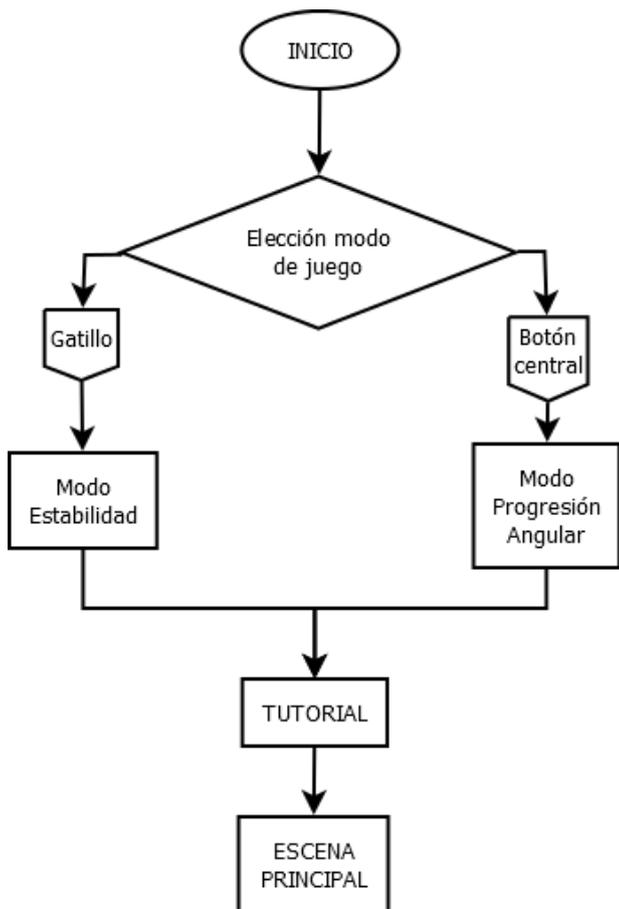


Fig. 5. Sucesión de escenas del videojuego

En la figura FFF se muestra el diagrama de flujo general del juego. En él se pueden ver las tres escenas de las que consta: escena inicial donde se elige el modo de juego, a continuación

se pasa al tutorial y finalmente a la escena de juego principal donde se desarrolla el juego en sí mismo. La elección del modo de juego la llevará a cabo el fisioterapeuta de acuerdo al paciente. En el caso de la plataforma de realidad virtual, se utilizará un mando que viene incluido con el visor de realidad virtual Oculus Go que mediante la pulsación de un gatillo o un botón el juego sabrá si se va a ejecutar un modo de juego u otro. Una vez seleccionado el modo de juego, éste se guarda como parámetro interno de manera que carga el tutorial adecuado a este modo y de la misma manera carga la escena de juego correspondiente a este modo de juego, ya que como anteriormente se describió existen notables diferencias entre ambos modos acorde a sus propios objetivos.

El juego ha sido desarrollado mediante el motor de desarrollo Unity y se ha utilizado C# como lenguaje de programación para la generación de los comportamientos de cada uno de los componente así como de todo conjunto del juego.

## VI. VALIDACIÓN Y PRUEBAS

Para validar el prototipo y hacer las pruebas pertinentes, se realizaron tres visitas a la clínica de fisioterapia del cliente. En cada una de ellas se hizo una demostración por parte del autor y seguidamente el propio fisioterapeuta probó el prototipo simulando las características de posibles pacientes. Después del cliente el prototipo fue probado por otras personas y al final de cada sesión se recogía la impresión del cliente así como las posibles mejoras que serían interesantes incorporar al prototipo. Cabe destacar la última sesión en la que un paciente real de la clínica estuvo probando el videojuego por lo que se hizo una primera aproximación a la funcionalidad real del dispositivo. Al término del proyecto el cliente generó un documento de valoración oficial en el que calificó el proyecto. Además se le entregó una encuesta de satisfacción en la que la valoración fue siempre el máximo para cada pregunta de la encuesta.



Fig. 6. Pruebas con un paciente real de la clínica de fisioterapia

## VII. CONCLUSIONES

La realidad virtual es una tecnología que puede tener mucho impacto en una gran variedad de ámbitos. El mundo de la rehabilitación no es ajeno a ello y mediante el desarrollo de un prototipo de acuerdo a una serie de especificaciones por parte de un fisioterapeuta se ha intentado probar que efectivamente el uso de esta tecnología puede tener un impacto muy positivo. Con este prototipo se ha conseguido que se puedan hacer los movimientos requeridos por un paciente mientras se encuentra abstraído en el mundo virtual, y esto ha generado un segundo punto positivo: la motivación para realizar los ejercicios. Al estar jugando, el paciente está entretenido, pensando en otro objetivo e inconscientemente está asumiendo los movimientos que tiene que realizar. Tampoco es un secreto que una persona motivada desarrolla mejor su potencial y hace las tareas con más efectividad, por lo que se puede concluir que el uso de videojuegos de realidad virtual en este ámbito es algo positivo que está despertando interés y que necesita de mayor desarrollo para obtener un mayor impacto positivo así como la necesidad de realizar estudios clínicos avalando el uso de esta tecnología en este campo.

## AGRADECIMIENTOS

Agradezco el apoyo de mi familia y amigos y en especial quiero dar las gracias a Miguel Ángel y a José, que nos ha ayudado mucho con su opinión y experiencia.

## REFERENCIAS

- [1] T. Prushansky y Z. Dvir, "Cervical Motion Testing: Methodology and Clinical Implications", vol. 31, 2008, pp. 503-508
- [2] P. Fernández-González, M. Carratalá-Tejada, E. Monge-Pereira, S. Collado-Vázquez, P. Sánchez-Herrera Baeza, A. Cuesta-Gómez, E. D. Oña-Simbaña, A. Jardón-Huete, F. Molina-Rueda, C. Balaguer-Bernaldo De Quirós, J. C. Miangolarra-Page y R. Cano-De La Cuerda, "Leap motion controlled video game-based therapy for upper limb rehabilitation in patients with Parkinson's disease: A feasibility study", *Journal of NeuroEngineering and Rehabilitation*, vol. 16, n° 1, 6 11 2019.
- [3] Y. Tokuyama, R. Rajapakse, S. Yamabe, K. Konno y Y.-P. Hung, "A kinect-based augmented reality game for lower limb exercise", *Proceedings - 2019 International Conference on Cyberworlds, CW 2019*, 2019.
- [4] C. Claucich, L. Carrere y C. Tabernig, "Virtual Reality Interface Built Using Unity3D for Rehabilitation with BCI Systems Based on Motor Imagery", *2018 IEEE Biennial Congress of Argentina, ARGENCON 2018*, 2019.
- [5] G. Riva, R. M. Baños, C. Botella, F. Mantovani y A. Gaggioli, "Transforming Experience: The Potential of Augmented Reality and Virtual Reality for Enhancing Personal and Clinical Change", *Frontiers in psychiatry*, vol. 7, p. 164, 30 9 2016.
- [6] D. Im, J. Ku, Y. Kim, S. Cho, Y. Cho, T. Lim, H. Lee, H. Kim y Y. J. Kang, "Utility of a Three-Dimensional Interactive Augmented Reality Program for Balance and Mobility Rehabilitation in the Elderly: A Feasibility Study", *Annals of rehabilitation medicine*, vol. 39, pp. 462-472, 1 6 2015.
- [7] H. Mousavi Hondori, M. Khademi, L. Dodakian, A. McKenzie, C. V. Lopes y S. C. Cramer, "Choice of human-computer interaction mode in stroke rehabilitation", *Neurorehabilitation and Neural Repair*, vol. 30, n° 3, pp. 258-265, 1 3 2016.
- [8] P. Kiper, A. Baba, M. Alhelou, G. Pregolato, L. Maistrello, M. Agostini y A. Turolla, "Assessment of the cervical spine mobility by immersive and non-immersive virtual reality", *Journal of Electromyography and Kinesiology*, vol. 51, p. , 2020.



# Análisis del tráfico Skype: Evaluación objetiva y subjetiva para diferentes condiciones de red

Autor: César Senés Romo, e-mail: cesarsenes@gmail.com

Tutor: Jorge Navarro Ortiz, e-mail: jorgenavarro@ugr.es

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—En la actualidad, las comunicaciones de voz sobre IP cada vez han tomado más importancia, por lo que han aparecido numerosas aplicaciones para realizar este tipo de comunicación. Entre ellas destaca *Skype* con millones de usuarios en todo el mundo.

Por este motivo, este trabajo fin de grado se centra en el desarrollo de una herramienta capaz de evaluar, en diferentes condiciones de red, el funcionamiento de *Skype*. El estudio realizado se basa en el análisis de la calidad de servicio (QoS) y calidad de experiencia (QoE) durante una llamada de voz con *Skype*.

Para la realización de este estudio, es necesario obtener el tráfico generado por ambos extremos de la comunicación. Procesando esta información, es posible caracterizar de forma objetiva la llamada realizada (e.g. calculando *throughput*, paquetes perdidos, retardo y *jitter*). Por otro lado, es necesario almacenar el audio utilizado durante la llamada en ambos extremos para su posterior procesamiento. Comparando el audio enviado con el recibido, se puede obtener una caracterización subjetiva de la llamada (e.g. mediante PESQ). Gracias a la herramienta desarrollada, se pueden introducir perturbaciones en la red para así observar cómo se comporta *Skype* en esas condiciones.

Finalmente, usando la aplicación desarrollada se han estudiado y presentado los resultados de varios escenarios. Estos escenarios se consideran relevantes para entender el funcionamiento de *Skype*, permitiendo ver cómo evolucionan la generación de tráfico y las estadísticas objetivas y subjetivas en función de diferentes condiciones de red. En particular, se ha analizado cómo afectan, por separado, las variaciones de *throughput*, del retardo y de las pérdidas de paquetes.

**Palabras clave**—Análisis objetivo, análisis subjetivo, calidad de experiencia (QoE), calidad de servicio (QoS), llamada de voz, *Skype*.

## I. INTRODUCCIÓN

ESTE proyecto se va a centrar en *Skype*, una de las aplicaciones de VoIP más usadas. Los últimos datos facilitados por *Microsoft*, en Febrero de 2015, nos indican que esta aplicación la utilizan 300 millones de usuarios activos mensualmente, tanto para entorno empresarial como para uso doméstico.

La fortaleza del códec de audio SILK es la utilización de *bitrate* variable, es decir, es capaz de ajustar la tasa de bits y frecuencia de muestreo dependiendo de las condiciones de la red. Como se muestra en el Cuadro I.

Debido a la importancia de *Skype* resulta interesante la implementación de una aplicación que permita estudiar detalladamente el funcionamiento de una llamada realizada a través de esta plataforma, así como la robustez de su códec

Modo	Bite Rate(Kbps)	Sample Rate(KHz)
Banda estrecha	6-20	8
Banda media	7-25	8,12
Banda ancha	3-30	8,12,16
Banda súper ancha	12-40	8,12,16,24

Tabla I  
MODOS OPERACIÓN SILK

en diferentes escenarios. Esto permitirá estudiar cómo afectan las diferentes condiciones de red tanto en los parámetros de QoS como de QoE en el usuario final.

Existen numerosas aplicaciones que monitorizan servicios VoIP pero. Sin embargo específicamente de *Skype*, el número es más reducido, debido a que el protocolo que usa *Skype* para las llamadas es cerrado y privativo. Por este motivo, las aplicaciones que analizan llamadas VoIP genéricas no suelen funcionar con *Skype*.

Por tanto el principal objetivo de este proyecto es el diseño de una aplicación cliente-servidor que permita evaluar de una forma objetiva y subjetiva la calidad de las llamadas a través de *Skype*.

Para este fin, durante la llamada, la aplicación captura los paquetes UDP generados por la llamada tanto en el cliente como en el servidor. Además se obtendrán las señales de audio generadas durante la llamada, tanto la original como la degradada por el efecto de la red. Con estos datos y mediante un post-procesado se calcularán las principales estadísticas, que se representarán gráficamente para caracterizar la calidad de la llamada.

## II. ESTADO DEL ARTE

Tras realizar un análisis de las aplicaciones y estudios que, al igual que este proyecto, se centran en la caracterización de llamadas realizadas mediante tecnología de VoIP, podemos concluir que ninguna de ellas cumple los objetivos propuestos en este trabajo. Resumiendo estos objetivos son, realizar un análisis tanto de la calidad de servicio como de experiencia durante una llamada de *Skype* con diferentes condiciones de red.

## III. IMPLEMENTACIÓN

En la Figura 1 se muestran un esquema en el cual se detalla secuencialmente como funciona la aplicación, con el fin de obtener una idea global de como interactúan cliente y servidor.

A continuación se presenta una breve explicación de este diagrama:

- 1) **Inicio Sesión en el cliente:** el usuario que actúa como cliente introduce en la aplicación diseñada un usuario y una contraseña, que este registrado previamente en Skype.
- 2) **Inicio Sesión en el servidor:** al igual que en el cliente, es necesario introducir un usuario y contraseña válidos, es decir registrados en *Skype*.
- 3) **Envío datos para la simulación al servidor:** desde el cliente se envía al servidor la información referente a la simulación. La información enviada principalmente es el audio que se reproduce durante la llamada y los perfiles de tráfico. Estos perfiles definen los parámetros introducidos al emulador de red. Y por último, el usuario de *Skype*, usado por el cliente, con el objeto de que el servidor pueda realizar la llamada. Esta información ha sido introducida por el cliente usando la interfaz gráfica de la aplicación creada.
- 4) **Inicio de la llamada y captura de información en el servidor:** usando el usuario recibido en el paso anterior, el servidor realiza una llamada, a través de la aplicación *Skype*. Por otro lado, se comienza a capturar, usando un analizador de protocolos, el tráfico generado durante la llamada. También simultáneamente se graba el audio de la llamada. Esta información será indispensable para el análisis y muestra de resultados final.
- 5) **Respuesta a la llamada entrante y captura de información en el cliente:** en este punto el cliente responde la llamada realizada por el servidor. Y del mismo modo que en el punto anterior se comienza a capturar el tráfico generado durante la llamada.
- 6) **Pérdidas en el enlace:** desde el servidor, se ejecuta el emulador de red. Éste irá variando los parámetros que caracterizan la red, dependiendo de lo introducido por el cliente al inicio de la simulación.
- 7) **Reproducción del audio de test durante la llamada:** una vez establecida la llamada desde el lado del cliente (sin pérdidas), se reproduce el audio elegido al principio.
- 8) **Fin de la llamada y envío de los datos recogidos en el servidor:** una vez que todos los eventos de la simulación se han completado, el servidor finaliza la llamada. Por último el servidor envía tanto la captura del tráfico y la grabación de la llamada al cliente.
- 9) **Análisis de los datos de los datos recogidos:** Finalmente, en el cliente se procede a analizar los resultados obtenidos. Para ello, se comparan las capturas del servidor (con pérdidas, introducidas por el emulador de red) y las del cliente (sin pérdidas). Con esta información es posible realizar un estudio de la calidad de servicio objetiva. Mientras que para obtener la calidad de servicio subjetiva, es necesario comparar el audio del test original con el grabado en el servidor (con pérdidas).

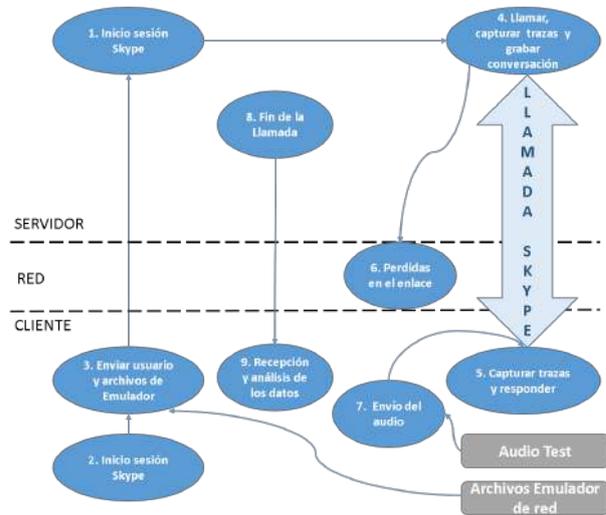


Fig. 1. Esquema de funcionamiento de la aplicación

usando máquinas virtuales. Más concretamente se ha usado el programa *VirtualBox*, mediante el cual se han creado dos equipos con sistema operativo *Windows 7*. Uno de ellos actúa como cliente mientras que el otro actúa como servidor.

Indicar que para conseguir enviar un audio desde el cliente al servidor a través de Skype, como se muestra en la Figura 2.

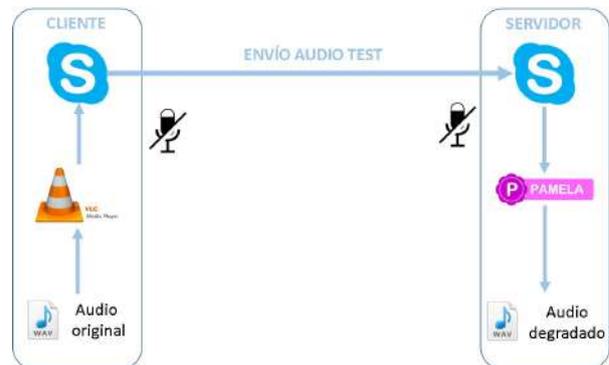


Fig. 2. Esquema de la configuración del sistema de audio.

Es necesario realizar una serie de modificaciones así como la utilización de programas de terceros en ambas máquinas virtuales.

#### IV. ESCENARIOS REALIZADOS Y RESULTADOS OBTENIDOS

En esta sección se muestran los resultados de varios escenarios. Los escenarios elegidos son representativos ya que muestran claramente como funciona el codec de *bitrate* variable de *Skype*. Para cada simulación, se muestran tanto estadísticas objetivas como subjetivas con el fin de realizar un análisis lo mas detallado posible. Para el modelado de estos escenarios se usa un audio de unos 600 seg (10 min). Este audio es el recomendado por la UIT (Unión Internacional de Telecomunicaciones) en la recomendación P.501 para la medición de calidad de transmisión telefónica.

Se ha desarrollado un programa en el lenguaje *JAVA*, que realiza todas las operaciones indicadas en la figura 1. Para conseguir la funcionalidad completa del programa, ha sido necesario el uso de aplicaciones de terceros.

Destacar que con el objeto de que este sistema sea portable y se pueda utilizar sin necesidad de realizar ningún tipo de configuración de red, el sistema se ha implementado

### A. Variando el Throughput

La planificación de la simulación realizada se muestra en la tabla II. Esta figura muestra, para cada instante de tiempo de la simulación, el perfil usado. En este caso, el perfil únicamente modifica la velocidad del enlace. Indicar que durante los primeros cinco segundos no actúa el emulador de red con el objetivo de que la llamada se establezca correctamente. A partir de ese instante cada 30 segundos va cambiando de manera progresiva el perfil.

En primer lugar, se muestran las gráficas en la Figura 3 que muestran la evolución del *throughput*, tanto en paquetes por segundo como en kilobits por segundo. En dicha figura se puede observar cómo, a pesar de que la velocidad del enlace es mayor a la mostrada, la tasa máxima de generación de paquetes para una llamada de Skype está entorno a 55 kbps en este caso. Como es de esperar, al bajar la velocidad del enlace, la generación de paquetes de Skype también disminuye.

El estudio realizado para esta simulación se divide en dos partes dependiendo de la objetividad de los resultados obtenidos.

Tabla II  
ESCENARIO DONDE EL *Throughput* ES VARIABLE.

Instante inicial	Perfil
0	sincronización
5	200kbps
35	150kbps
65	100kbs
95	75kbps
125	50kbps
155	25kbps
185	10kbps
215	25kbps
245	50kbps
275	75kbps
305	100kbps
335	150kbps
365	200kbps
395	150kbps
425	100kbps
455	75kbps
485	50kbps
515	25kbps
581	-

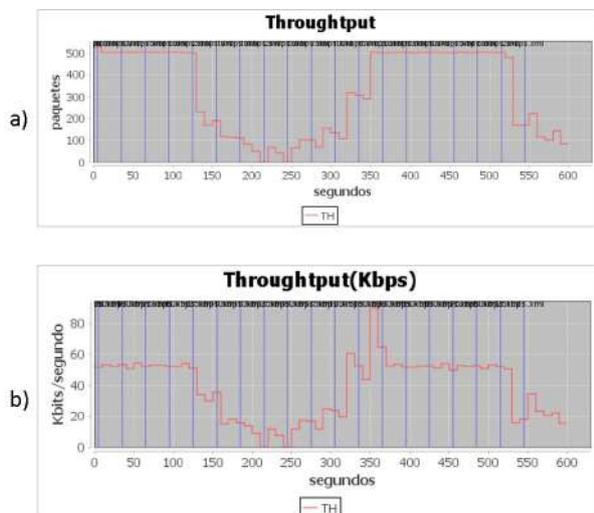


Fig. 3. Gráfica evolución del *Throughput* durante la simulación.

En primer lugar se puede observar las gráficas objetivas. Se muestra la tasa de paquetes perdidos, tanto en número de paquetes perdidos 4 a) como en porcentaje 4 b). Se observa cómo, cuando la velocidad de Skype es igual a la máxima (55 kbps), la pérdida de paquetes es nula. Pero cuando la velocidad disminuye por debajo de este valor, aparecen paquetes perdidos. Este efecto se debe al tiempo que tarda Skype en adaptar su *bitrate*. Es decir, en los cambios de perfil de red (en el servidor), el cliente continúa emitiendo a la velocidad anterior hasta que descubre que ahora las condiciones de red son otras. Hasta ese instante de tiempo, el cliente emite más información de la que puede recibir el servidor (debido a las restricciones impuestas por el emulador de red). Por este motivo se produce la pérdida de paquetes.

A continuación, se muestran las gráficas relacionadas con el retardo de la simulación. Se observa cómo, al disminuir la velocidad del enlace, aparece retardo en los paquetes. Esto, como en el caso anterior, se debe a que el cliente emite a mayor velocidad que lo que se puede transmitir por la red hacia el servidor (cuando se produce un cambio en las condiciones de la red). Los paquetes que no pueden ser recibidos en el servidor se almacenan en una cola del emulador de red, a la espera de que puedan ser enviados, ocasionando así un retardo mayor.

Finalmente, se muestra las gráficas relacionadas con el *jitter*. En estas gráficas también aparecen efectos de *jitter* cuando el cliente modifica su tasa de generación de paquetes, ocasionando que el tiempo entre paquetes no sea siempre constante y aumentando así el *jitter*.

A continuación se muestra las gráficas Subjetivas. En primer lugar, se muestra la gráfica con la información de *Perceptual Evaluation of Speech Quality*, PESQ. En esta gráfica, al comparar ambas señales de voz al bajar la velocidad del enlace, empeora la calidad de la llamada.

Finalmente se muestra la información relacionada con el E-Model.

Como conclusión general, en este caso se ve claramente cómo, al bajar la velocidad del enlace de la red por debajo de 30 kbps, la calidad de la llamada realizada se ve directamente afectada. Este resultado era de esperar, ya que la función E-Model depende de las estadísticas objetivas de retardo y de tasa de paquetes perdidos.

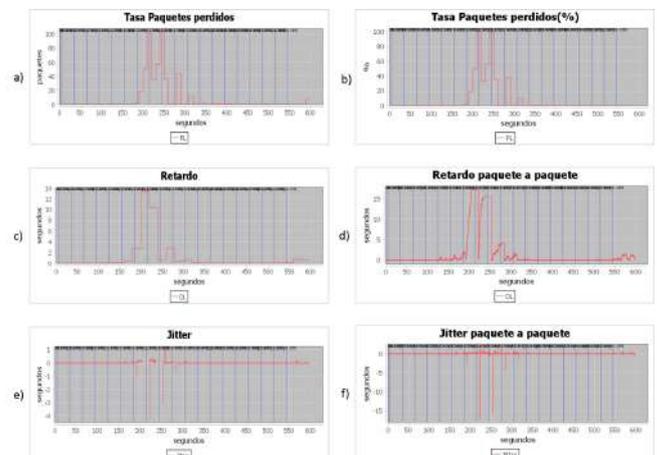


Fig. 4. Gráficas objetivas para la simulación con *throughput* variable.

**B. Variando la tasa de paquetes perdidos**

En el cuadro III se muestra cómo va ir evolucionando la simulación a lo largo del tiempo. Durante la simulación solo se introducen pérdidas por parte del emulador, no se introduce ninguna perturbación más.

En primer lugar, se muestra la Figura 6 que incluye las gráficas relacionadas con la pérdida de paquetes.

Destacar que la Figura 6 a) muestra el porcentaje de paquetes perdidos. Dado que el emulador no introduce otro tipo de pérdidas, este porcentaje coincide en cada instante de tiempo con el mostrado en el Cuadro III.

Por otro lado, la gráfica mostrada 6 b) muestra los paquetes perdidos. Se puede observar cómo, al aumentar el porcentaje de pérdidas, disminuyen los paquetes perdidos. Esto se debe a que el cliente detecta que la red tiene pérdidas por lo que usa un codec con el que se mandan menos paquetes, lo que implica que se pierdan también menos paquetes.

El estudio realizado para esta simulación se divide en dos partes dependiendo de la objetividad de los resultados obtenidos.

Ahora se muestran en la Figura 7 las gráficas con las estadísticas objetivas de la simulación.

En primer lugar, observando 7 a) y b), al detectar *Skype* que en el enlace hay pérdidas comienza a enviar a una velocidad menor, seguramente con el objetivo de que se congestione menos la red y se pierdan el menor número posible de paquetes.

Con respecto al retardo y al *jitter*, se observan pequeñas fluctuaciones poco significativas. Estas fluctuaciones es probable que aparezcan debido al cambio de bitrate de la llamada.

A continuación, se muestra en la Figura 5 las gráficas con las estadísticas subjetivas de la simulación.

En este caso se puede observar cómo ambas gráficas mantienen una forma similar. Estos gráficos indican que, al comenzar la simulación, la comunicación era muy buena pero, conforme han ido aumentando el porcentaje de paquetes perdidos, la calidad de la llamada se ha deteriorado. Se comprueba así que las pérdidas afectan de forma similar tanto al PESQ como al E-Model.

**C. Variando el retardo**

En el cuadro IV se muestra cómo va ir evolucionando la simulación a lo largo del tiempo. Los perfiles de tráfico

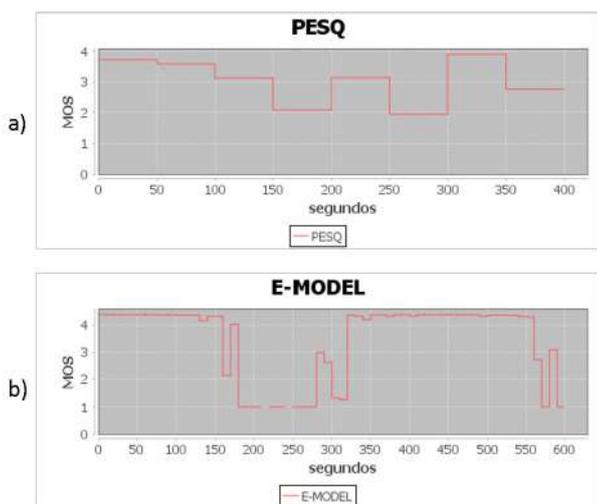


Fig. 5. Gráficas subjetivas para la simulación con *throughput* variable.

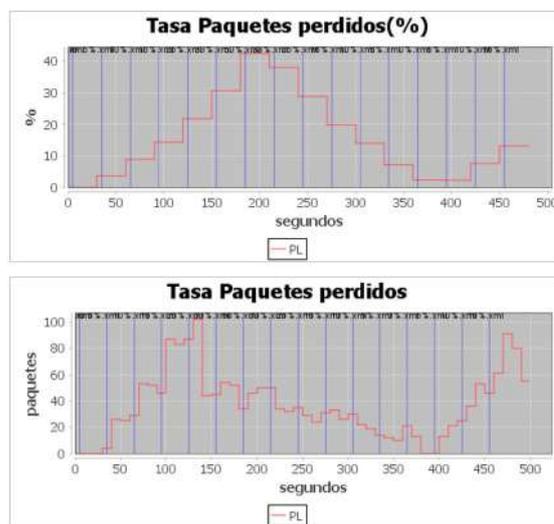


Fig. 6. Gráfica evolución de la tasa de paquetes perdidos durante la simulación.

Tabla III  
ESCENARIO DONDE LA TASA DE PAQUETES PERDIDOS ES VARIABLE.

Instante inicial	Perfil
0	sincronización
5	0%
35	5%
65	10%
95	16%
125	25%
155	33.3%
185	50%
215	33.3%
245	25%
275	16%
305	10%
335	5%
365	0%
395	5%
425	10%
455	16%
500	-

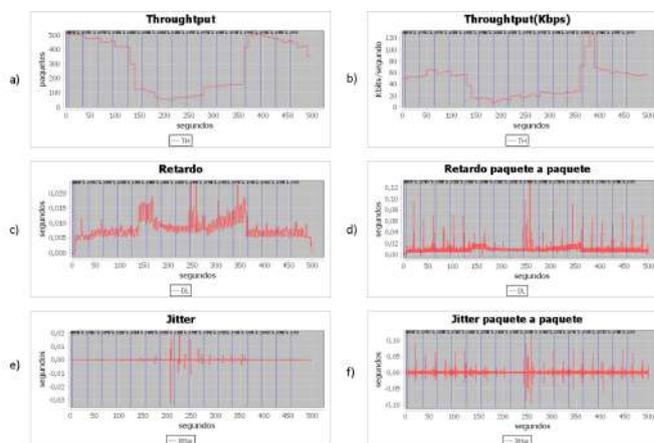


Fig. 7. Gráficas objetivas para la simulación con tasa de paquetes perdidos variable.

mostrados en la tabla indican los milisegundos que va a tardar un paquete en llegar de un extremo a otro.

La Figura 9 muestra la evolución del retardo durante la simulación.

El estudio realizado para esta simulación se divide en

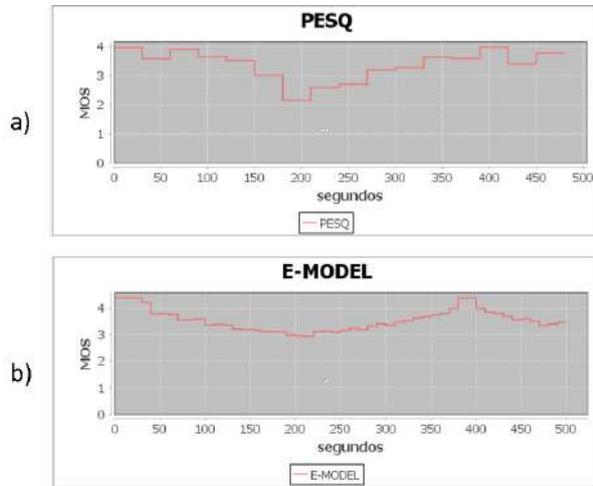


Fig. 8. Gráficas subjetivas para la simulación con tasa de paquetes perdidos variable.

Tabla IV  
ESCENARIO DONDE EL RETARDO ES VARIABLE.

Instante inicial	Perfil
0	sincronización
5	0 ms
35	50 ms
65	100 ms
95	200 ms
125	300 ms
155	500 ms
185	750 ms
215	1000 ms
245	750 ms
275	500 ms
305	300 ms
335	200 ms
365	100 ms
395	50 ms
425	0 ms
500	-

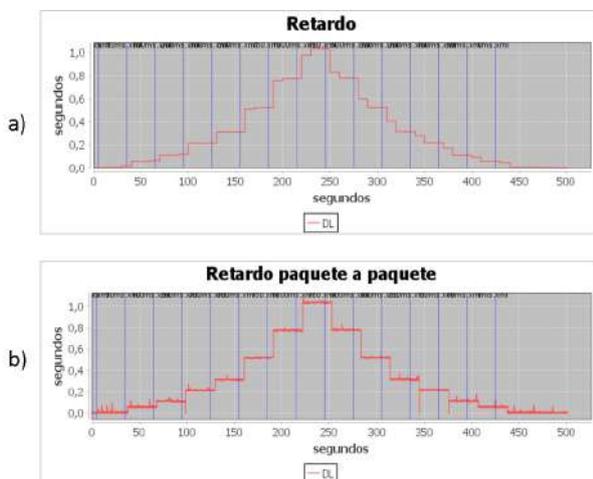


Fig. 9. Gráfica evolución del retardo durante la simulación

dos partes dependiendo de la objetividad de los resultados obtenidos.

Ahora se muestran en la 10 las gráficas con las estadísticas objetivas de la simulación.

Observando la Figura 10 a) y b), como ha ocurrido en las otras simulaciones, al detectar el cliente que algo va mal, la cantidad de información enviada del cliente al servidor disminuye drásticamente.

Por otro lado observando la Figura 10 c) y d), se detecta que se pierden algunos paquetes. A mayor retardo, mayor número de paquetes perdidos, si bien las pérdidas no son demasiado elevadas.

Finalmente, se observa que el *jitter* permanece principalmente constante, a excepción de algunas pequeñas fluctuaciones.

A continuación se muestra en la Figura 11 las gráficas con las estadísticas subjetivas de la simulación.

En este caso entre las dos gráficas de E-Model y PESQ son muy diferentes. Esto se debe a que E-Model se calcula usando las estadísticas objetivas, por lo que al aumentar el retardo disminuye el MOS según el E-model.

Sin embargo, para calcular el MOS según PESQ se utilizan los audios capturados en ambos extremos. PESQ realiza operaciones de sincronización entre ambos audios para poder

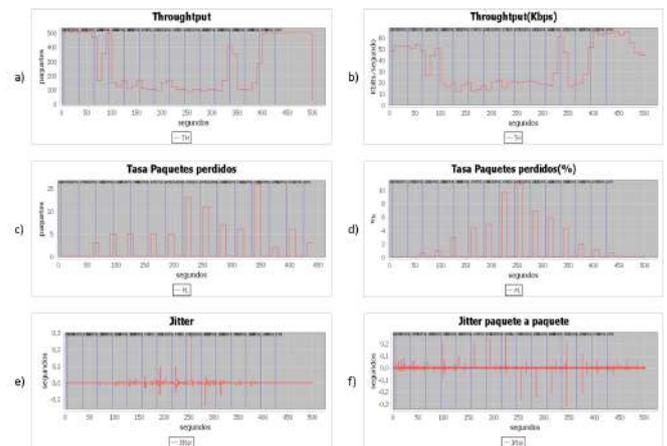


Fig. 10. Gráficas objetivas para la simulación con retardo variable.

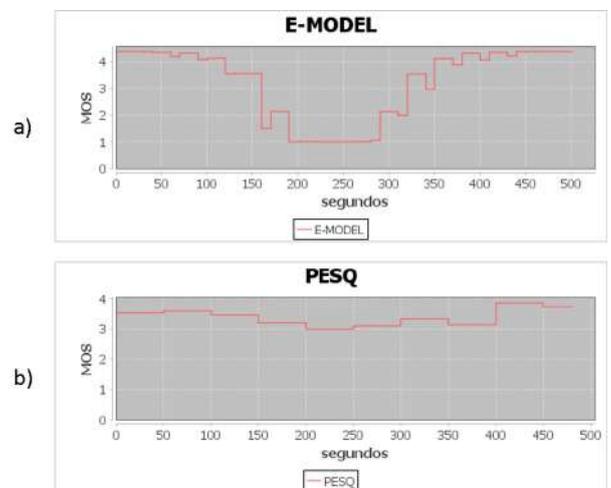


Fig. 11. Gráficas subjetivas para la simulación con retardo variable.

comparar ambas señales, la original y la recibida tras atravesar la red. Esta sincronización hace que el retardo de los paquetes no afecte a los valores calculados por PESQ, que permanece prácticamente constante. Las pequeñas variaciones seguramente se deben a los paquetes perdidos al introducir retardos elevados.

## V. CONCLUSIONES Y LÍNEAS FUTURAS

Como conclusión principal, la solución realizada consigue cumplir con todos los requisitos iniciales para este trabajo. Ya que se ha conseguido realizar una herramienta capaz de analizar el tráfico generado por *Skype*. Tal como se comprueba con la realización del estado del arte, no se ha encontrado ninguna aplicación que tenga esta funcionalidad.

La principal contribución que aporta este proyecto es la disposición de una herramienta altamente configurable con la que poder evaluar un gran número de escenarios y observar cómo se comporta *Skype* en estas situaciones.

Pese a que cumple con todas las especificaciones iniciales, como trabajos futuros se presentan una serie de posibles mejoras:

- Mejorar la interfaz con el usuario, haciéndola más atractiva y más funcional para el usuario.
- Instalar las aplicaciones en dos ordenadores diferentes, ya que para este trabajo se han usado dos máquinas virtuales. Sería interesante hacerlas funcionar en dos ordenadores dentro o fuera de una misma red realizando NAT (suponiendo que usen direcciones IP privadas) y realizar test en redes reales.
- Adaptar la herramienta desarrollada para que sea capaz de funcionar en otros sistemas operativos.
- Realizar más tests, ya que sería interesante mezclar distintos tipos de pérdidas y retardos para comprender pormenorizadamente el funcionamiento de *Skype*.
- Añadir más gráficas subjetivas. En este proyecto se han incluido dos medidas subjetivas, pero sería interesante implementar alguna más.
- Para que el estudio subjetivo sea más completo, sería interesante realizar tests con usuarios reales y recoger la experiencia del usuario.

## AGRADECIMIENTOS

Quisiera agradecer a mi familia por el apoyo recibido en todo momento, a mi padre y a mi hermano por estar a mi lado siempre que los he necesitado y en especial a mi madre que también ha vivido intensamente este proyecto, preguntando todos los días y preocupándose por la evolución.

También quiero agradecer a mi tutor Jorge Navarro por el apoyo recibido en todo momento, sin duda sin su ayuda hubiera sido mucho más complicado conseguir lograr los objetivos de este proyecto.

Finalmente a Almudena que me ha ayudado con su compañía y apoyo a que estos últimos meses hayan sido mucho más agradables.

# IMPLEMENTACIÓN DE UN ROBOT CONTROLADO DESDE UN DISPOSITIVO MÓVIL

Autor: Javier Montijano Valdivia, e-mail: javimoval@correo.ugr.es

Tutor: Juan José Ramos Muñoz; e-mail: jjramos@ugr.es

Tutor: Jonathan Prados Garzón; e-mail: jpg@ugr.es

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación  
**Departamento de Teoría de la Señal, Telemática y Comunicaciones**  
**Universidad de Granada**

**Resumen**— En la actualidad existen diversas plataformas que permiten programar robots en lenguajes de alto nivel, pero para poder personalizar la capa de red y que el presupuesto sea de bajo coste, nos centraremos en placas de desarrollo como la Raspberry Pi [1]. Parte de la motivación de este proyecto es que sirva como precedente para el desarrollo de otros robots y plataformas por estudiantes de la Universidad.

En este proyecto se ha desarrollado un robot similar a un coche controlable telemáticamente utilizando el paradigma cliente servidor, mediante Websockets [2] para la comunicación entre ellos. El servidor está programado en Java [3] y el cliente en HTML y Javascript [4] con la intención de que se adapte a distintos dispositivos. Con este fin, ha sido necesario estudiar el funcionamiento del protocolo usado en los Websockets, así como la plataforma Raspberry Pi.

Se trata de un proyecto que funciona en distintas plataformas, se han integrado distintas librerías, tanto de la parte de red como del uso físico de los de las entradas y salidas hardware de la plataforma Raspberry, interconectando distintas tecnologías y descubriendo la versatilidad y sencillez que nos ofrece el protocolo Websockets.

**Palabras clave**— móvil, java, javascript, raspberry pi, robot, websockets.

## I. INTRODUCCIÓN

La posibilidad de crear plataformas y dispositivos visibles a raíz de los conocimientos impartidos en el Grado es, en sí mismo, un incentivo para entusiasmarse y motivarse. Además, en ámbito personal, he ido desarrollando una inquietud a desarrollar e implementar ciertos servicios de distinta índole, que sean capaces de comunicarse entre sí y que faciliten su uso con tecnologías comunes y de uso doméstico, como es por ejemplo el uso de Smartphones.

En la actualidad existen ya diversas plataformas sencillas para la programación de movimientos de un robot, para usuarios más y menos avanzados. Aunque, si queremos desarrollar la parte de programación en red con un protocolo propio, tendremos que ahondar en plataformas como las que presentaremos a continuación.

Por tanto, nos disponemos a crear una plataforma versátil con un protocolo sencillo y piezas de bajo coste que sean fácilmente replicables, cuyo propósito sea que se use esta

plataforma o modificaciones de la misma para motivar el estudio y desarrollo de las competencias propias de este proyecto, y a su vez que nos permita diseñar un protocolo e implementar el mismo de manera real, con lenguajes de alto nivel, como pueden ser Java.

## II. OBJETIVOS

En este trabajo fin de grado se desarrolla una plataforma robótica de bajo coste que sea programable y controlable telemáticamente. El robot, que debe ser de bajo coste, podrá ser programado en algún lenguaje de alto nivel.

En resumen, los objetivos del proyecto son:

1. Desarrollar una plataforma hardware robotizada.
2. Diseñar e implementar la plataforma software correspondiente para programar el robot.
3. Desarrollar como ejemplo una aplicación móvil para controlar el robot.

## III. ESTADO DEL ARTE

Hoy día existen ya en el mercado diversas plataformas preparadas para programar los movimientos de un robot mediante lenguajes de alto nivel, como podrían ser Edison y Lego Mindstorm. Cumpliendo con los objetivos de nuestro proyecto, buscamos una plataforma que nos permita una mayor personalización, hasta el punto de poder controlar la programación de la capa de red.

Por lo que, analizamos plataformas existentes que permitan la programación de tarjetas en lenguajes de alto nivel, con capacidad de interactuar con sensores y actuadores estándar y comunes, en búsqueda de un proyecto de bajo coste. En el desarrollo de la memoria analizamos las diferencias y semejanzas de las placas Arduino Uno, Orange Pi Zero y Raspberry Pi

Existen muchas más placas de desarrollo con estas características y posibilidades, nos hemos centrado en las más populares y relevantes. Fijarnos en éstas nos permitirá poder documentar y conseguir bibliografía más fácilmente ante los nuevos retos y problemáticas, debido a su amplio uso en multitud de proyectos.

Con la elección final de Raspberry Pi, utilizaremos su sistema operativo Raspbian [5], una distribución de Linux

modificada especialmente para estas placas ARM. En sus repositorios encontraremos ya software totalmente compatible con Raspberry Pi que hará más sencilla la implementación de nuevos servicios, y centrándonos en la implementación de los servicios propios a desarrollar. Utilizar Raspbian y todo el desarrollo existente en torno a él, es un factor clave más para la elección de esta plataforma.

#### IV. DISEÑO

Con el desarrollo del proyecto, y con el objetivo de programar un cliente que funcionara en dispositivos comunes, descubrimos el Protocolo WebSockets, que nos permite implementarlo en cualquier navegador web moderno, tanto en ordenador como en dispositivos móviles. Su diseño está ideado para funcionar en navegadores y servidores web, y se puede utilizar por cualquier aplicación cliente/servidor. Su uso supone una gran ventaja, y es que el software específico que necesita el cliente para comunicarse con el servidor es un navegador web, por lo que cualquier ordenador y smartphone podrá utilizar este protocolo sin ningún requisito especial, sin tener que caer en la casuística de programación para crear una "app" para cada tipo de lenguaje de cada dispositivo compatible, ya que HTML, Javascript y Websockets son entendidos por la inmensa mayoría de estos dispositivos antes mencionados.

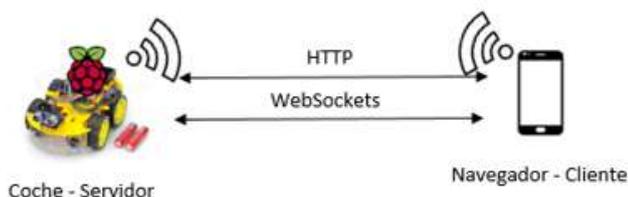


Fig. 1. Entorno de desarrollo del proyecto

El robot y proyecto funcionan con el paradigma de cliente – servidor, tal como aparece en la figura 1.

El servidor, integrado en el coche (robot que integra la placa Raspberry Pi), funciona mediante 5 clases distintas, cuyo propósito es dividir las partes funcionales del código para su modificación, cambios, etc. Su cometido es el control de la GPIO de la placa según las órdenes del cliente, atendiendo y respondiendo solicitudes del protocolo Websockets. También se encarga de controlar las distancias de los objetos en su delantera para evitar un choque, tomando decisiones frente a las órdenes del cliente. Además, para el correcto funcionamiento del sistema, el servidor crea un punto de acceso Wifi, esto nos permite utilizar el coche sin necesidad de disponer otro punto de acceso y configurar el coche para que se conecte a él, además así evitamos alguna restricción de los cortafuegos de otro punto de acceso o router. [6]

Con la elección de Raspberry Pi, utilizaremos su sistema operativo Raspbian, una distribución de Linux modificada especialmente para estas placas ARM. En sus repositorios encontraremos ya software totalmente compatible con Raspberry Pi que hará más sencilla la implementación de nuevos servicios, y centrándonos en la implementación de los servicios propios a desarrollar. Utilizar Raspbian y todo el

desarrollo existente en torno a él, es un factor clave más para la elección de esta plataforma.

Por otra parte, el cliente, ejecutado en cualquier navegador web compatible con websockets, simplemente tiene que conectarse al punto inalámbrico que crea el servidor, acceder a la web "localhost" y tendrá acceso al mando a distancia. Este mando servirá para mandar órdenes de movimientos al robot, desde simples movimientos hasta controlando la velocidad. A su vez, en la misma pantalla del mando, se muestra información de forma visualmente clara del estado actual del robot, como el movimiento que está haciendo en ese momento, si se ha parado porque iba a producirse un choque, etc.

Para la comunicación entre cliente y servidor se ha desarrollado un sencillo protocolo llamado "CarPi", como ejemplo del desarrollo de este robot, y se añaden instrucciones para la creación correcta del protocolo según el robot que finalmente se desarrolle utilizando este hardware y software.

#### V. IMPLEMENTACIÓN

A continuación, se relatan los detalles de la implementación del proyecto, haciendo hincapié en algunos métodos utilizados que se consideran más importantes. Además, se añaden algunos diagramas de estados de las hebras que manejan el coche.

##### A. Servidor

El funcionamiento del programa en el servidor se basa en tres hebras que funcionan simultáneamente: ServidorCoche, Interprete y ControlChoque, además de otras 2 clases que sirven de apoyo para estas hebras y sus labores, como son Coche y Ws.

La hebra ServidorCoche, y hebra principal, ejecuta el resto de hebras y se encarga de la comunicación y eventos relacionados con el canal de comunicación. Prepara el canal de comunicación a la espera de solicitudes, avisa cuando se conecta un nuevo cliente, recibe los mensajes y los añade a la cola, interpreta si el mensaje que recibe es una solicitud de parada y lo procesa de manera prioritaria, avisa del cierre de conexión, etc.

La hebra Interprete está a la espera revisando la cola de mensajes por si encuentra alguno nuevo, cuando encuentra alguno nuevo lo interpreta y procesa con los métodos de la clase Coche, después de ejecutar las órdenes en Coche, duerme durante el tiempo que se solicita en el mensaje para que se realice el movimiento en ese tiempo.

La hebra ControlChoque se encarga de toda la gestión del sensor de ultrasonidos, a nivel de la configuración de la GPIO, así como del cálculo de distancias, de la generación de un evento cuando hay peligro de colisión, y de la solicitud a la clase Coche de un frenado de emergencia para que no se produzca el choque.

La clase Coche posee las constantes del patillaje GPIO que se conecta a los motores, así como los métodos necesarios para configurar las salidas GPIO para los movimientos disponibles.

Esta clase incorpora algunos métodos nuevos y sobrescribe otros de la librería base de Websockets que importamos que es `org.java_websocket`, extiende la misma.

### B. Cliente

El único archivo programado para el cliente es el archivo `index.html`, que se observa su ejecución en la figura 2. Este archivo lo solicita y recibe el cliente al conectar con el servidor apache del coche. Se trata de una página web con una interfaz responsiva e intuitiva, que indica al usuario información sobre el estado de sus órdenes al servidor y sobre el servidor en sí.

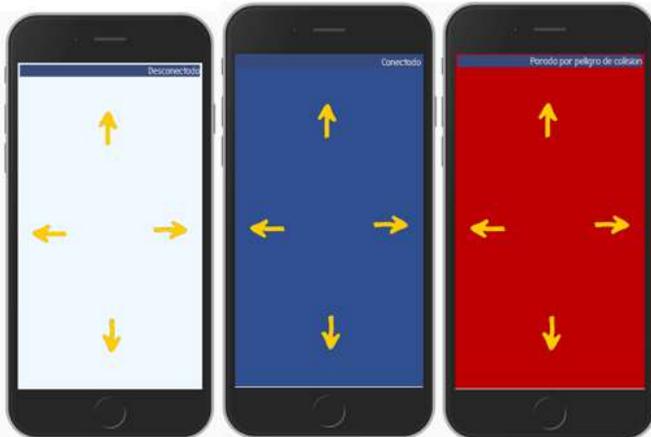


Fig. 2. Interfaz del cliente. Distintos estados Desconectado – Conectado – Colisión

Contiene fundamentalmente dos partes, un pequeño fragmento con código HTML, que crea una tabla de 3 filas y 3 columnas donde coloca las flechas de dirección, también crea una franja en la parte superior de la web que muestra información sobre el coche.

A su vez, utilizamos algunas configuraciones CSS [7], incorporadas en este mismo código, con la intención de hacer una web responsiva, esto quiere decir, que se adapte a los distintos dispositivos que ejecuten la web.

Esta parte del código es totalmente estática, pero necesaria para mostrar la web, ya que el código Javascript modificará algunas partes de este código estático y capturará acciones en este código, como por ejemplo cuando tocamos una flecha.

La mayoría del código de esta web es Javascript, y se encarga de dos funcionalidades primordialmente. La primera es capturar los toques del usuario en la web, por lo que el código captará cuando se hace click en una flecha, o cuando se levanta el dedo, o cuando se arrastra por la pantalla. Según cada una de estas acciones y la localización de las mismas en la página web ejecuta unas funciones que se encargan de comunicar con el servidor la acción deseada según lo que se haya hecho en la web. Por último, este parte del código también se encarga de mostrar avisos en la franja superior de la web y de cambiar el color de fondo de la misma web según las respuestas del servidor y sus estados.

## VI. CONCLUSIONES

En el mercado existen una amplia gama de robots cuyos movimientos y acciones son más sencillas de programar en lenguajes de alto nivel que nuestra solución propuesta. El inconveniente de éstos es que son menos personalizables y tienen un coste mayor.

La solución que hemos planteado nos permite modificar la mayoría de la parte software del coche, además de que es lo suficientemente potente para ejecutar Websockets como servidor, un servidor web apache, gestionar una red wifi, con un servidor DHCP [8] incluido. Esto permitirá en las asignaturas relacionadas con el diseño e implementación de protocolos de red, disponer de una plataforma asequible y versátil para desarrollar prácticas de la asignatura.

El uso de Raspberry y Raspbian ha sido una ventaja para encontrar librerías y API's adecuadas a nuestro proyecto, así como para resolver los problemas de programación gracias a la gran comunidad de desarrolladores que publica sus avances en la red de sus avances personales.

Aunque para elegir el protocolo de comunicación de red hayamos pasado por un largo tiempo de pruebas, la elección final por usar Websockets ha simplificado notablemente el código que debe implementar el desarrollador tanto en el servidor como en el cliente. Además, que esté escrito en Javascript y lo interpreten la mayoría de los navegadores modernos nos permite desarrollar interfaces para clientes compatibles con multitud de dispositivos pues los requisitos de los mismos son muy pequeños.

Después de desarrollar un proyecto de esta índole y resolver los problemas más comunes en este tipo de desarrollos, se vislumbra más sencillo embarcarse en nuevos proyectos con robots que se comunican por red y tienen otros cometidos distintos.

Concluir también, que la formación recibida en las asignaturas de programación y aquellas que han tenido relación con el desarrollo de aplicaciones en red han sido una gran ayuda para solucionar los pequeños retos que van surgiendo en la implementación del código.

## VII. TRABAJO FUTURO

Tras desarrollar y documentar un proyecto así, se nos ocurren varios por los que podrían continuar desarrollando un trabajo futuro.

El servidor está preparado para funcionar con una cola de mensajes, pero en este momento del desarrollo el mando no está preparado para enviar una serie de mensajes de forma que el coche los procese uno a uno sin intervención humana. Por lo que modificar la interfaz del cliente para activar una opción y poder grabar una secuencia de movimientos según el tiempo que se mantenga pulsado cada botón y al final enviar toda la secuencia sería un desarrollo interesante.

La estructura del código, en cuanto a paradas por solicitud del cliente, o por obstáculos, se ha ido incorporando conforme se han incorporado las funciones y se han realizado pruebas en los mismos. Debido a estas pruebas algunas la clase de webservice llama directamente a las funciones de la clase coche, como la clase ControlChoque. Aunque llegamos a esta decisión al evaluar la eficacia de estas funciones para evitar choques, estos cambios con el cometido de evitar choques del coche y bloqueos de las hebras, hacen que entender y leer el código sea más complicado, por lo que otra mejora para un futuro desarrollo sería modificar el código para que las clases tuvieran unas funciones más marcadas respetando la urgencia de los eventos de paradas.

El código Javascript se ejecuta localmente en el navegador del cliente, y aunque existen unos estándares, para hacer una web responsiva para todos los dispositivos hay que tener en cuenta particularidades de configuración que se deben establecer para distintos terminales del cliente. En este caso se ha configurado para terminales con iOS. Como proyecto futuro, sería interesante testarlo en terminales con otros sistemas y configurar la página web del cliente para que no produzca fallos en otros sistemas.

Dadas estas pautas de funcionamiento y conexión, sería interesante incluir más funcionalidades en este robot, como más sensores, alguna pantalla que indicara estados. Incluso sin muchas modificaciones, se podría incluir un sistema que hiciera al coche moverse solo por un espacio sin chocar con los obstáculos gracias a su sensor de ultrasonidos.

#### AGRADECIMIENTOS

Gracias a los tutores por su esfuerzo y paciencia para llevar a término este proyecto, a la ETSIIT y el Grado en Ingeniería de Tecnologías de Telecomunicación por la formación recibida que hacen que sea posible llevar a cabo este proyecto.

#### REFERENCIAS

- [1] Mocq, François – ‘Raspberry Pi 2 Utilice todo el potencial de su nano-ordenador’ 2016. Barcelona: Ediciones ENI
- [2] IETF, The WebSocket Protocol <https://tools.ietf.org/html/rfc6455>
- [3] Johan Vos, Java API for WebSocket <http://www.oracle.com/technetwork/articles/java/jsr356-1937161.html>
- [4] Christian Vigouroux, “Aprender a desarrollar con JavaScript” Ediciones ENI
- [5] Harrington, William ‘Learning Raspbian: get up and running with Raspbian and make the most out of your Raspberry Pi’ Birmingham, UK
- [6] Jouni Malinen, ‘hostapd: IEEE 802.11 AP, IEEE 802.1X / WPA / WPA2 / EAP / RADIUS Authenticator’ <http://w1.fi/hostapd/>
- [7] W3C - Guía Breve de CSS <http://www.w3c.es/Divulgacion/GuiasBreves/HojasEstilo>
- [8] DNetwork Working Group, Dynamic Host Configuration Protocol <https://www.ietf.org/rfc/rfc2131.txt>

# LoRaWAN Gateway and IoT Low-Cost Mote Prototype

Autor: Ángel Guzmán Martínez, e-mail: anguzmar@correo.ugr.es

Tutor: Jorge Navarro Ortiz, e-mail: jorgenavarro@ugr.es

Tutor: Sandra Sendra Compte, e-mail: ssendra@ugr.es

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—Hoy en día el Internet de las Cosas o *Internet of Things* es un término que está cobrando cada vez más relevancia. Debido al crecimiento de los dispositivos pertenecientes al Internet de las Cosas, las Redes de Baja Potencia o *Low Power Wide Area Networks* están también en auge, ya que suplen algunas de las carencias propias de las redes tradicionales, como WiFi o las redes celulares.

Existen varios tipos de Redes de Baja Potencia, pero en este proyecto nos centramos en la tecnología de red LoRaWAN o *Long Range Wide Area Network*. En consecuencia, en este proyecto hemos desarrollado dos dispositivos relacionados con LoRaWAN: una mota de bajo coste que actuará como transmisor y un gateway donde recibiremos los datos.

Por último, es importante hacer énfasis en que la mota debe ser de bajo coste, ya que uno de los objetivos de este proyecto es probar que el uso de LoRaWAN no requiere una fuerte inversión económica.

**Palabras clave**—Gateway, Internet de las Cosas, LoRaWAN, mota, Redes de Baja Potencia.

## I. INTRODUCCIÓN

ESTE primer capítulo nos sirve de punto de partida, donde se presentan los siguientes conceptos:

- 1) Alcance y objetivos del proyecto. Los cuales quedan englobados como el desarrollo de una mota de bajo coste y un gateway LoRaWAN de forma que pueden comunicarse entre ellos.
- 2) Motivación y contexto. Estos dos conceptos giran entorno al auge del Internet de las Cosas y la necesidad de usar Redes de Baja Potencia.
- 3) Estructura de la memoria.

Además, para dar una mejor perspectiva a la motivación, se hace una breve reseña a la situación de LoRaWAN en España, Fig. 1.

## II. INTRODUCCIÓN A LORAWAN

LoRaWAN es relativamente reciente por lo que este segundo capítulo nos ayudará a tener un mejor entendimiento de esta tecnología.

Sin embargo, LoRaWAN se basa en la modulación de capa física LoRa. De forma que, en primer lugar, tenemos que profundizar sobre cómo funciona dicha modulación, con la ayuda de la especificación sobre LoRa proporcionada por Semtech [1]. Los puntos más significativos de la modulación LoRa son:

- Uso de una señal *chirp* como señal moduladora, Fig. 2.
- Adaptabilidad. Permite cambiar la velocidad de transmisión en función de la distancia que deseamos cubrir.
- Robustez frente a interferencias.

Una vez vista la modulación LoRa, pasamos a introducir



Fig. 1: Gateways LoRaWAN registradas en España en junio de 2017.

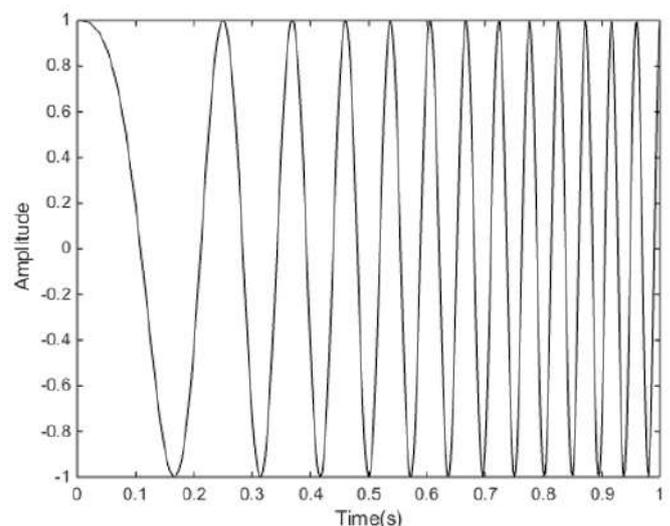


Fig. 2: Señal Chirp.

los fundamentos básicos de LoRaWAN que nos permitirán el desarrollo de este proyecto. Estos fundamentos provienen de la especificación LoRaWAN definida por LoRa Alliance [2]. Los conceptos que se ven en esta sección son los listados a continuación:

- Topología. Cómo viaja un paquete y por que elementos pasa a través de una red LoRaWAN.
- Clase de dispositivos. Según cómo se realiza la comunicación podemos distinguir entre tres clases de dispositivos: clases A, B y C.
- Parámetros regionales. Necesarios para conocer la normativa que se aplica en las comunicaciones LoRaWAN de nuestra zona.
- Seguridad a nivel de paquete: integridad y confidencialidad.
- TTN o *The Things Network*. TTN es una red LoRaWAN pública que usaremos debido a que no disponemos de nuestro propio servidor de red LoRaWAN.

### III. ESTADO DEL ARTE

Antes de comenzar la elaboración del proyecto, debemos hacer un estudio sobre las diferentes posibilidades que tenemos para llevarlo a cabo. Dicho estudio incluye las ventajas y desventajas de cada opción, así como una estimación del coste que conllevaría. Debido a la naturaleza de este proyecto, esta sección se divide en dos subsecciones, una para la mota y otra para el gateway.

#### A. Gateway

Para el desarrollo del gateway se consideraron principalmente tres opciones:

- 1) Monocanal. No cumple totalmente con las especificaciones de LoRaWAN, pero es la opción más barata. Coste estimado: 51.06€.
- 2) Multicanal *Do It Yourself*. Esta opción cumple con las especificaciones LoRaWAN pero requiere montaje tanto *hardware* como *software* por nuestra parte. Coste estimado: 243.30€. Esta es la opción que elegimos para llevar a cabo el proyecto.
- 3) Multicanal ya montada. Cumple con las especificaciones y no requiere apenas montaje, a cambio de ser más cara. Coste estimado: 300€.

#### B. Mota

En cuanto a la mota, hay una gran variedad de opciones, pero se pueden agrupar en dos grupos:

- Mota con la capa LoRaWAN MAC ya implementada. Este tipo de motas son más caras, a cambio de requerir menos tiempo para el montaje.
- Mota sin la capa LoRaWAN MAC implementada.

Teniendo en cuenta que uno de los objetivos del proyecto consiste en minimizar el coste de la mota, hemos optado por una mota que no tiene la capa LoRaWAN MAC implementada. Normalmente, esta opción implicaría tener que implementar LoRaWAN por nosotros mismos. Afortunadamente, existen librerías con las funciones de la capa MAC que necesitamos ya implementadas. Dichas librerías son uno de los principales factores que permiten minimizar el coste de la mota.

### IV. PLANIFICACIÓN Y ESTIMACIÓN DEL COSTE

Considerando el tipo de gateway y mota elegidos en el capítulo anterior, este capítulo busca hacer una estimación tanto temporal como económica del coste del proyecto.

La planificación temporal se puede resumir con un diagrama de Gantt, Fig. 3.

La estimación del coste total del proyecto se encuentra en la Tabla I.

### V. ESPECIFICACIÓN DE REQUISITOS

Antes de pasar al diseño de los dispositivos, tenemos que definir las características y prestaciones que queremos conseguir en el producto final.

Estos requisitos están agrupados en funcionales y no funcionales y engloban campos tales como la seguridad, la comunicación, la eficiencia energética y el coste, entre otros.

Por lo tanto, el diseño de nuestros dispositivos estará condicionado por los requisitos especificados en este capítulo.

### VI. DISEÑO

Este capítulo se divide en dos partes, una primera parte dedicada al diseño de la mota y otra para el gateway.

#### A. Mota

Como hemos mencionado anteriormente, la mota va a estar basada en un chip que no tenga la capa LoRaWAN MAC implementada. Es por esto que nos hemos fijado en el chip ESP12. Este chip es en realidad un chip WiFi, pero podemos reprogramarlo para que funcione con LoRa. En consecuencia, hemos optado por comprar la placa de desarrollo Wemos D1 mini, Fig. 4, que contiene el chip ESP12.

A continuación, necesitamos un modulador LoRa y una antena para transmitir nuestra señal. El dispositivo LoRa BEE de Dragino, Fig. 5, incluye tanto la antena como el modulador LoRa y es el que vamos a usar para nuestra mota.

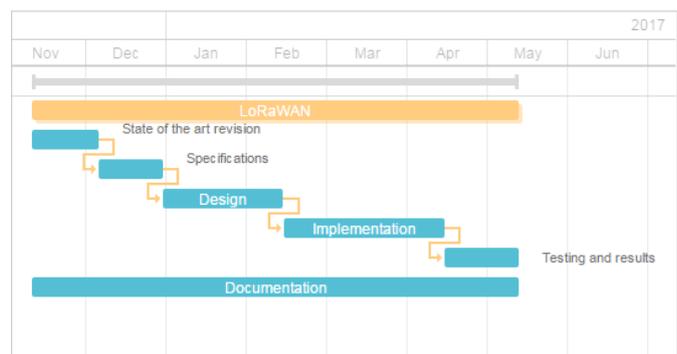


Fig. 3: Diagrama de Gantt.

Tabla I: Estimación del coste total del proyecto.

Concepto	Coste asociado
Recursos humanos	14.480€
Propósito general	87,74€
Mota	€26,9
Gateway	249,65€
<b>Total</b>	<b>14.844,29€</b>

## B. Gateway

El *gateway* va a estar compuesto principalmente de dos elementos:

- Una Raspberry Pi, que hará la función de centro neurálgico del *gateway*.
- Un concentrador. Es el componente que permite que nuestro *gateway* sea multicanal y cumpla con las especificaciones de LoRaWAN.

Además de estos dos componentes, también necesitaremos una antena, un *pigtail* y una caja metálica hermética. La caja es necesaria porque queremos que el *gateway* esté situado en el exterior para una mejor cobertura.

## VII. IMPLEMENTACIÓN

De nuevo, tenemos que dividir el capítulo en varias subsecciones. Pero, en este caso, lo vamos a dividir en tres:

- 1) Aplicación final.
- 2) Mota.
- 3) *Gateway*.

### A. Aplicación Final

En esta aplicación será donde recibiremos los paquetes transmitidos por la mota y los descriptaremos. Tenemos que empezar implementado la aplicación porque será aquí donde generemos las claves de seguridad, que tendremos que aplicar en la mota.



Fig. 4: Placa de desarrollo Wemos D1 Mini.



Fig. 5: Transceptor LoRa y antena LoRa BEE.

Como estamos usando la red TTN, registramos la aplicación en su página web, donde tenemos que especificar una serie de parámetros como el identificador y la región de nuestra mota.

### B. Mota

La implementación de la mota comienza por conectar la placa Wemos con el transceptor LoRa BEE, Fig. 6.

Una vez conectados los distintos elementos, pasamos a configurar el *software*. En nuestra caso, hemos programado la mota usando Arduino y la librería LMIC, en cuyo *sketch* hay que incluir ciertos parámetros como las claves de seguridad, las conexiones entre los pines, el identificador de la mota, etc.

Por último, cargamos el *sketch* de Arduino en la placa Wemos y ya debería empezar a transmitir paquetes.

### C. Gateway

Al igual que hicimos con la mota, el primer paso es hacer las conexiones oportunas entre el concentrador y la Raspberry Pi, Fig. 7.

En cuanto al *Software*, lo primero que tenemos que hacer es instalar un sistema operativo en la Raspberry Pi, nosotros hemos usado Raspbian Jessie Lite.

Con el sistema operativo ya instalado, tenemos que pasar a configurar los parámetros del *gateway*. Afortunadamente,

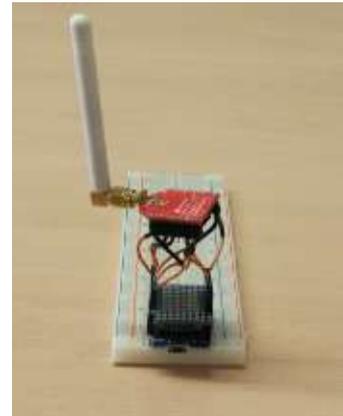


Fig. 6: Mota una vez hechas las conexiones entre sus componentes.



Fig. 7: Conexiones entre el concentrador y la Raspberry Pi.

TTN nos proporciona un *script* que facilita este proceso enormemente. De esta forma, ejecutamos el *script*, rellenamos una serie de campos como las coordenadas de nuestro *gateway* y listo.

El siguiente paso consiste en registrar el *gateway* en la página web de TTN, para que este pueda transmitir los paquetes a través de su red y recibirlos en la aplicación que creamos anteriormente.

Por último, después de asegurarnos que todo está correcto, cerramos la caja para evitar que entre agua o polvo y la ponemos en el lugar elevado en el exterior. Nosotros optamos por la terraza de la quinta planta de nuestra facultad.

### VIII. PRUEBAS Y ANÁLISIS DE LOS RESULTADOS

Con ambos dispositivos, la mota y el *gateway*, ya montados, nos queda someterlos a una serie de pruebas con el fin de comprobar que funcionan correctamente. Las pruebas realizadas fueron las siguientes:

- 1) Comprobación de que los paquetes se reciben correctamente en la aplicación.
- 2) Medición de la cobertura del *gateway*, tanto interior como exterior.
- 3) Captura de los saltos de canal durante la transmisión de paquetes.
- 4) Análisis del espectro de una transmisión ascendente.
- 5) Medición de la potencia consumida por la mota.

Todas las pruebas dieron buenos resultados, siendo los más relevantes:

- La recepción de los paquetes en la aplicación, Fig. 8. Hemos recibido con éxito, en hexadecimal, la cadena "Hello World" que nuestra mota está transmitiendo.
- El mapa de cobertura exterior, Fig. 9. En este mapa, el punto más lejano que medimos se encuentra a unos 550 metros del *gateway*.
- El consumo de potencia en un caso realista, Fig. 10. A partir de esta medida hemos calculado que el consumo sería, aproximadamente, 0.0406W.

### IX. CONCLUSIONES

Con las pruebas del capítulo anterior, queda demostrado el largo alcance de la modulación LoRa. También podemos ver como afecta la densidad de edificios, que reduce el alcance considerablemente. Tomando este mapa de cobertura como referencia, suponemos que podríamos sacarle el máximo partido a la cobertura en zonas rurales donde la densidad de edificios sea baja.

Así mismo, la prueba de consumo nos indica que esta tecnología podría ser muy útil para dispositivos que dependan



time	received	sent
11:08:22	26	1
11:08:30	25	1
11:40:38	24	1
11:47:44	23	1
11:47:34	22	1

Fig. 8: Paquetes recibidos con éxito en la aplicación.

de baterías. Esto daría lugar a una reducción del coste de mantenimiento, ya que minimizaría la frecuencia con la que tenemos que cambiar las baterías.

Por otro lado, hemos demostrado con éxito que, si disponemos de un *gateway* público cercano, el coste que implica usar LoRaWAN es especialmente bajo y asequible.

Finalmente, queda claro como LoRaWAN es una tecnología emergente con mucho potencial que podría, si se dan las circunstancias adecuadas, alcanzar la misma relevancia que WiFi o las redes celulares.

### AGRADECIMIENTOS

Agradecimiento especial a Jorge por haberse involucrado tanto con este proyecto. A pesar de ser el tutor de otros muchos proyectos y de su propio trabajo, siempre estaba dispuesto a echar un cable.

Mención especial también a Sandra, Miguel Ángel y Pablo por acoger a nuestro humilde *gateway* en la terraza de su despacho y actuar como nuestro equipo de mantenimiento personal.

Por último, mis amigos y familia que me estuvieron apoyando durante todo el año, así como presionándome para que acabase la memoria de una vez.

### REFERENCIAS

[1] Semtech Corporation, "LoRa Modulation Basics", mayo de 2015. Disponible: <http://www.semtech.com/images/datasheet/an1200.22.pdf>



Fig. 9: Mapa de cobertura exterior basado en medidas experimentales.



Fig. 10: Medida de la potencia consumida por la mota en un caso realista.

- [2] N. Sornin, M. Luis, T. Eirich, T. Kramp y O.Hersent, "LoRaWAN Specification", versión 1.02, julio de 2016.

**Ángel Guzmán Martínez** Nacido en Baza, Granada en 1993. Graduado en Ingeniería de Tecnologías de Telecomunicación, con mención en telemática, por la Universidad de Granada. Actualmente cursa el primer año del Máster de Ingeniería en Tecnologías de Telecomunicación en la Universidad de Granada.



# Security improvement for LoRaWAN networks

Author: Natalia Chinchilla Romero, e-mail: nataliachr@correo.ugr.es

Tutor: Jorge Navarro Ortiz, e-mail: jorgenavarro@ugr.es

Degree: Telecommunication Technologies Engineering

Department of Signal Theory, Telematics and Communications of the University of Granada  
University of Granada

**Abstract**—The Internet of Things is one of the hottest topics in communications today, with around 7 billion connected devices nowadays and an unprecedented growth expected for next years. LoRaWAN is one of the most outstanding technologies for massive IoT communications due to features such as large coverage and low power consumption. In this context, where end-devices may be placed on remote or unattended locations, physical security is a major concern. In this work, a cost-effective solution is proposed to provide hardware security to LoRaWAN end-devices. A 3GPP-like security scheme is designed and implemented as a proof-of-concept using real equipment. By using this solution, end-devices will be no longer vulnerable to security threats such as impersonation or cloning attacks.

**Index Terms**—LoRaWAN, 3GPP, security, USIM cards.

## I. INTRODUCTION

THE Internet Of Things (IoT) is one of the hottest topic in communications today. Current IoT devices vary from 6 to 9 billion devices (e.g. 7 billions in 2018 according to IoT Analytics [1]) and forecasts predict from 20 to 30 billion IoT devices by 2020 [2]. These values represent market revenues from \$151B in 2018 up to \$1,567B by 2025.

Many of the IoT services can be considered as massive Machine-Type Communications (mMTC), one of the three major 5G use cases. These services assume a massive number of battery-powered devices in a high number of locations and environments. For this reason, mMTC requirements are low-power communications and a wide range coverage.

The main technologies which fulfill these requirements are based on cellular networks, such as NB-IoT (Narrowband-IoT) which uses licensed frequency bands, or on Low Power Wide Area Network (LPWAN) technologies, such as SigFox or LoRaWAN (Long-Range Wide Area Network) that uses license-exempt frequency bands.

LPWAN technologies were born with IoT requirements from the very beginning, such as long range (up to several tens of kilometers), very low power consumption (years of battery operation) and very low bandwidth (tens of kbps). Due to its popularity, low cost and the possibility of setting up an own network, this work will be focused on LoRaWAN.

IoT devices are exposed to many security vulnerabilities due to their resource-constrained nature. Although the LoRaWAN standard implements several mechanisms to protect data in transit between the LoRaWAN core network entities and connected end-devices, the nodes can still be vulnerable to physical attacks. If stored keys can be extracted, for example, then a device can be impersonated on the network. In remote

areas that are not continuously monitored, such as rural areas, the vulnerability to physical attacks is very high.

In this work, it is proposed to use Universal Subscriber Identity Module (USIM) cards as hardware security modules for LoRaWAN end-devices. So that, the stored key cannot be extracted by e.g. reading the memory chips of the node. Also, USIM cards are very cheap and available to operators, so the economic impact of this solution will be negligible since, as it will be explained, no other circuitry will be needed. Additionally, in order to emulate current 3GPP Security, it is proposed to use the 128-bit ciphering and integrity keys CK and IK as the LoRaWAN application and network session keys (*AppSKey* and *NwkSKey*), and to implement the 3GPP Authentication and Key Agreement (AKA) signaling procedure between the LoRaWAN end-devices and the application server. Thus, this solution employs the 3GPP authentication for the initial attach and the LoRaWAN procedures for ciphering and integrity using the derived keys. As the authentication is only performed during the initial attach, the impact on the battery lifetime should be negligible. However, since keys may be stored on the device RAM memory during execution, it is recommended to update these session keys periodically (e.g. every day) in order to avoid device impersonation. As a proof-of-concept, the solution is implemented using real LoRaWAN equipment.

The content of this work has led to the sending of the article "Improving Hardware Security for LoRaWAN" to the *IEEE Conference on Standards for Communications and Networking (CSCN)*, October 2019, Granada (Spain).

## II. STATE OF THE ART

Although the literature about hardware security for Internet of Things (IoT) is abundant, there are only a few works focused on the Long-Range Wide Area Network (LoRaWAN) technology. In this section it is named some hardware related to communicate SIM/USIM cards with *Arduino* software. Also, it is named some works related to LoRaWAN security.

The *Arduino GSM* (Global System for Mobile Communications) shield [3] is a module which allows *Arduino* to connect to the Internet, receive SMS (Short Message Service) messages and make voice calls to *Arduino's GSM* (Global System for Mobile communications) library. This device allows communication with a SIM (Subscriber Identity Module) card using AT (*ATtention*) commands.

The SparkFun LTE CAT M1/NB-IoT Shield-SARA-R4 board is equipped with an Arduino compatible microcontroller [4]. It carries wireless cellular functionality and high bandwidth for IoT applications. Like LoRaWAN technology, it maintains low power consumption and small size. It has a modem called SARA-R410M-02B LTE CAT MTE / NB-IoT, which provides cellular communication to IoT devices through two LP-WAN technologies. This board works by introducing a USIM to allow Long Term Evolution (LTE) communications.

Jeon *et al.* propose in [5] the design of a secure element to be included in LoRaWAN devices, which safely stores keys and performs security functions such as integrity and ciphering. However, few details are given about implementation (no proof-of-concept is provided), cost or power consumption.

### III. LORAWAN SECURITY

LoRaWAN [6] is an open standard for Low Power Wide Area Network (LPWAN), which uses LoRa (Long-Range) [7] (property of Semtech) or Frequency Shift Keying (FSK) as physical layer. Its main features low-cost (in terms of infrastructure, operation expenditure and end-devices), interoperability (since it is standardized), low power (battery lifetime up to 20 years), long range (up to 30 miles in rural areas), GPS-free geolocation, security (thanks to end-to-end AES (Advanced Encryption Standard) encryption) and high capacity.

LoRaWAN is optimized for low power consumption and is designed to support large networks with millions of devices. Similarly, security is also designed for low power consumption, low implementation complexity, low cost and high scalability [8]. LoRaWAN achieves mutual authentication, integrity protection and confidentiality, as summarized in Fig. 1.

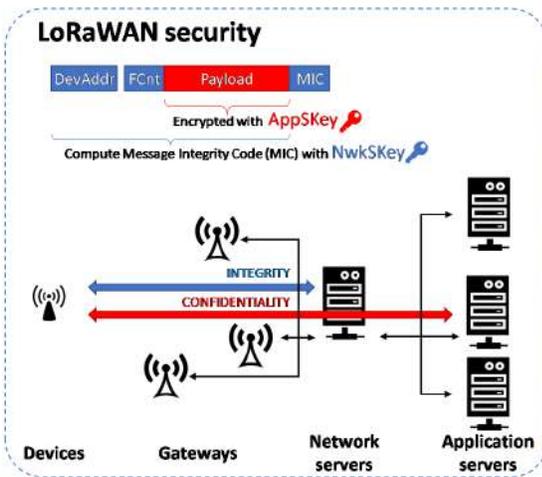


Fig. 1. LoRaWAN security.

LoRaWAN devices can be activated either via Over The Air Activation (OTAA) when a device is deployed or reset, or via Activation By Personalization (ABP) [6] (see Fig. 2). Independently of the activation method, the device stores a device address (DevAddr), an application identifier (AppEUI), a network session key (NwkSKey), and an application session

key (AppSKey). In the case of devices activated by personalization (ABP), these parameters are directly included into the device’s firmware. In the case of OTAA, these values are derived from a globally unique device identifier (DevEUI), an application identifier (AppEUI), and an AES-128 key (AppKey). In this case, the join request and accept messages—exchanged between the device and the network server—include a message integrity code (MIC) computed with the AppKey, which allows each end to verify that the other end knows this key, thus achieving mutual authentication.

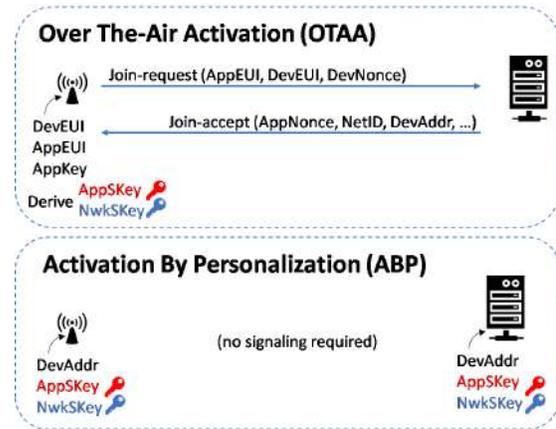


Fig. 2. LoRaWAN activation.

For data messages, integrity is achieved by also adding a MIC code which is computed using the NwkSKey. This field is used by both the end-device and the network server to verify data integrity. Finally, end-to-end encryption is performed for application payloads using an encryption scheme based on AES with the 128-bit application session key (AppSKey).

As commented, end-devices should store only necessary keys required for communication. Every device should hold unique, random keys such that a compromised end-device should not be able to compromise other end-devices. If these keys are obtained, an attacker would be able to produce legit, signed and encrypted messages. This makes essential to avoid physical attacks such as e.g. reading memory chips to capture this information.

### IV. 3GPP SECURITY

3GPP security [9] for 4G networks is based on the Authentication and Key Agreement (AKA) procedure, which allows both the User Equipment (UE) and the evolved Node B (eNB) to achieve mutual authentication and to generate session ciphering (CK) and integrity (IK) keys.

As shown in Fig. 3, when a UE connects to the Evolved Packet Core (EPC) over the Evolved Universal Terrestrial Radio Access Network (E-UTRAN), the AKA procedure is performed between the UE and the Mobility Management Entity (MME). The procedure starts when the UE tries to attach to the mobile network by sending an attach request with its identity. This request is retransmitted to the Home Subscriber Server (HSS), which generates the Authentication

Vector (AV). This AV contains a random number (RAND), an authentication token (AUTN), the ciphering and integrity keys (CK, IK) and the expected response (XRES). The AV is sent to the MME, which sends an authentication request to the UE with the RAND and the AUTN parameters. With this information, the Universal Subscriber Identity Module (USIM) is able to compute the response (RES), which is forwarded to the MME. If RES matches XRES, the UE is considered authenticated and the ciphering algorithm is then negotiated.

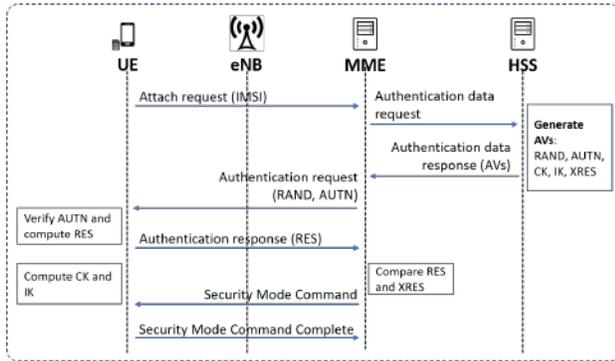


Fig. 3. 3GPP security (EPS AKA).

## V. DESIGN PROPOSAL

What is proposed in this work is to enhance LoRaWAN security by adding 3GPP Security with USIM cards installed into the end-devices. This idea includes a simple hardware design proposal and the creation of a 3GPP Security Server which interacts with the LoRaWAN application server.

### A. LoRaWAN device communication with USIM card

The design assumes that it is able to directly communicate the end-device with a USIM card using e.g. the SCLib library [10]. This library allows the communication with a smart card directly connected to e.g. an Arduino board, for both asynchronous and synchronous smart cards and using T0 protocol (T1 partially supported).

USIM cards are composed of a microprocessor, different types of memories and logic to manage security. The filesystem is stored in an internal EEPROM and has a hierarchical structure with a root called Master File (MF). Within this filesystem, there are two types of files: Dedicated Files (DFs) and Elementary Files (EFs). A file identifier (ID) is used to address or identify each specific file. The ID consists of two bytes and is encoded in hexadecimal notation. A DF contains only a *header* (i.e. acts as a directory), whereas an EF contains a *header* and a *body* (i.e. acts as a file). The header contains meta-information (e.g. available space under DF, number of EFs and children DFs, length of a record, access conditions to an EF, etcetera). The body contains information to the application of the smart card. In a USIM card, there are three types of EF: *transparent EF*, *linear fixed EF* and *cyclic EF*. The first one is organized as a sequence of bytes. The second one is organized in records, which are groups of bytes with a known

coding. The last one is organized in records but implementing a circular buffer, i.e. there are no first and last records but previous and next ones.

The operations to get access to the filesystem are coded into a set of commands. These commands are sent using APDUs (Application Protocol Data Units). A command APDU sent to the card contains a mandatory 4-byte header (CLA (Instruction Class), INS (Instruction Code), P1 (Instruction Parameter 1), P2 (Instruction Parameter 2)) and from 0 to 65 535 bytes of data. The response APDU sent by the card contains from 0 to 65 536 bytes of data, and 2 mandatory status bytes (SW1 (Status Word 1) and SW2 (Status Word 2)).

The main commands are described next:

- **SELECT**: selects a file and makes its header available.
- **GET RESPONSE**: gets data after a command.
- **READ BINARY**: reads the body of a transparent EF.
- **READ RECORD**: reads one record of a linear fixed or cyclic EF.
- **VERIFY, CHANGE, ENABLE, DISABLE, UNBLOCK CHV**: permits the management of *Personal Identification Number 1* (PIN1) and 2 (PIN2).

The main APDUs utilized for authentication in a USIM are summarized in Table I, which follows the appropriate order to communicate with the USIM card.

TABLE I  
APDUS FOR USIM AUTHENTICATION

SELECT FILE $EF_{DIR}$						
CLA	INS	P1	P2	Lc	Data	Le
00	A4	08	04	02	2F00	00
GET RESPONSE						
00	C0	00	00			
READ RECORDS						
00	B2	01	04	26	XX	
SELECT FILE $ADF_{USIM}$						
00	A4	04	04			
INTERNAL AUTHENTICATE						
00	88	00	81	22	XX	00

According to ETSI TS 102 221 [11], the MF contains five EFs which are mandatory for 3GPP. One of them is the  $EF_{DIR}$  file, which contains the Application Identifiers (AIDs) and the Application Labels of the 3GPP applications present on the card as mandatory elements. After selecting this EF, a successful response contains SW1=61 and SW2 indicates the number of response bytes still available. The response, read as a record, includes a pointer to the Application DF  $ADF_{USIM}$ . After selecting the  $ADF_{USIM}$ , the authenticate command is performed with RAND (User Authentication Challenge) and AUTN (Authentication Token for Network) as parameters. If the response starts with 'DB', the 3G authentication was successful and RES (Authentication Response), CK (Ciphering Key), and IK (Integrity Key) are returned. If the response starts with 'DC', there was a synchronization failure and the AUTS (Authentication Token for Re-Synchronization) parameter is returned, which allows the network to compute the correct SQN (Sequence Number) (by using AK (Anonymity Key))

in order to recover synchronization with the terminal. If re-synchronization is required, the process is repeated with a new AUTN calculated with the new SQN.

USIM cards consist of a series of physical contacts, which includes power supply (C1), reset (C2), clock (C3), ground (C5), programming voltage (C6) and input/output data (C7).

Based on the SCLib library, it has been implemented the USIMAuth library for the Arduino platform, which has been tested on the Arduino Mega 2560 [12] and the ESP32 [13] boards. Algorithm 1 depicts the usage of the developed library.

The provisions of contacts that it has been used for each board are shown in table II.

**Algorithm 1** Usage of the *USIMAuth* library

```

usimauth.init(C1pin, C2pin, C3pin, C7pin);
usimauth.perform3GAuth(RAND, AUTN);
if (usimauth.are3GKeysAvailable()) {
    RES=usimauth.getRES();
    CK=usimauth.getCK();
    IK=usimauth.getIK();
} else if (usimauth.isAUTSAvailable()) {
    AUTS=usimauth.getAUTS();
}
    
```

TABLE II  
PROVISION OF PINS

PIN Arduino Mega 2560	PIN ESP32 board	Contact USIM card
12	17	C1 - Vcc
7	13	C2 - RST
11	21	C3 - CLK
GND	GND	C5 - GND
10	12	C7 - I/O

**B. 3GPP-like signaling between LoRaWAN device and servers**

A 3GPP Security Server to emulate the Evolved Packet System (EPS) AKA procedure (see Section IV) between LoRaWAN end-devices and their application server has been designed.

The security server interacts with the LoRaWAN application server in order to 1) read (and filter) the messages sent from the end-devices, 2) send messages to the end-devices, and 3) to modify application server parameters such as the session keys.

The messages exchanged between each of the LoRaWAN entities are exposed in Fig. 4. The 3GPP AKA procedure is implemented as a protocol state machine. The state diagram is shown in Fig. 5 and explained next.

- **State 1: LoRaWAN initialization:** For simplicity, it is assumed ABP, although the activation method is not relevant since the information exchanged does not require ciphering. Then, the end-device sends an “Attach Request (DevAddr)” to the security server and receives an “Authentication Request 1 (RAND)”, where RAND is a 128-bit random number.

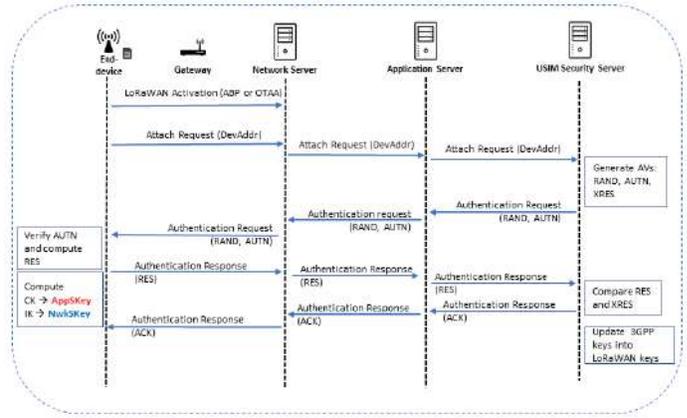


Fig. 4. Mapping of LoRaWAN-3GPP communication.

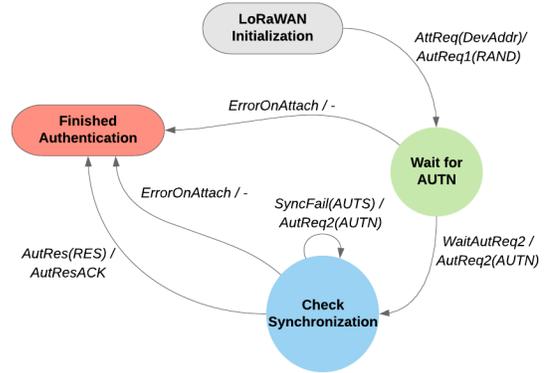


Fig. 5. 3GPP AKA protocol state machine.

- **State 2: Wait for AUTN:** Due to the downlink limitations of LoRaWAN, it has been split the “Authentication Request (RAND, AUTN)” message sent in the EPS AKA procedure into two messages: the previous “Authentication Request 1 (RAND)” and a new “Authentication Request 2 (AUTN)” message. In this state, the end-device sends a “Wait for Authentication Request 2” message and waits for the reception of a “Authentication Request 2” message containing the AUTN parameter.
- **State 3: Check synchronization:** After receiving the AUTN, the end-device executes the 3G authentication procedure. If the USIM card is synchronized (checked by computing the SQN from the received AUTN), the end-device sends an “Authentication Response (RES)” message and waits for the reception of an “Authentication Response ACK”. Otherwise, the end-device sends a “Synchronization Failed (AUTS)” message. In that case, the 3GPP Security Server computes a new AUTN (with the SQN extracted from the AUTS), and sends a new “Authentication Request 2 (AUTN)” message.
- **State 4: Finished Authentication:** After a successful authentication or upon a failure (signaled by an “Error on Attach” message), the procedure is considered complete. If the authentication was not successful, the end-device

will need to restart again the whole procedure in order to authenticate.

## VI. PROOF OF CONCEPT

To carry out a proof of concept, it has been implemented the prototype shown in Fig. 6. The end-device with the USIM is connected to the laptop using a USB cable in order to see the logs sent from the serial port. These logs include LoRaWAN frames exchanged between the end-device and the gateway, as well as the communication with the USIM card.

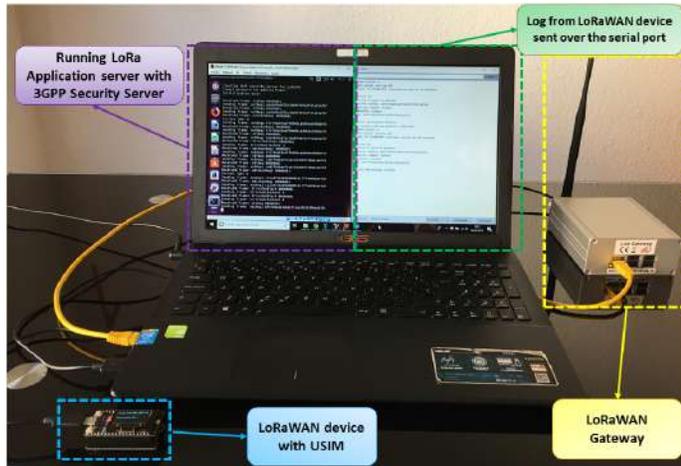


Fig. 6. Proof-of-concept prototype.

The LoRaWAN gateway is an IMST's Lite Gateway [14], which is based on a Raspberry Pi B+ and an ic880A LoRaWAN concentrator for the ISM (Industrial, Scientific and Medical) 868 MHz band. It is directly connected via Ethernet to the laptop which executes the LoRaWAN servers. The gateway implements the Packet Forwarder from the The Things Network Zurich's ic880a-gateway project [15], which is responsible for forwarding data packets between the end-device and the LoRa Network Server.

The LoRaWAN servers are implemented using the Lora Server project [16]. The components of this project (LoRa Gateway Bridge, LoRa App Server, MQTT Broker, Redis and PostgreSQL databases) and the new 3GPP Security Server designed are virtualized and executed in a laptop with an Intel i7-4510U processor and 4GB of RAM. The host operating system is the 64-bit version of MS Windows 10 Home, which executes VirtualBox 6.0 with an Ubuntu 16.04 guest OS.

The 3GPP Security Server interacts with the LoRaWAN application server by means of the MQTT protocol and the HTTP RESTful API of the LoRa Server project. Our security server subscribes to the "application+/device+/rx" topic in order to read (and filter) the messages sent from the end-devices. Similarly, the RESTful API allows the security server to send messages to the end-devices, but also to modify application server parameters such as the session keys.

For the management of end-devices and their USIM cards, the security server includes a PostgreSQL database called "lora\_usim", which contains the following fields: *DevEUI*,

*DevAddr*, *CurrentProtocolState*, *usim\_key*, *usim\_op*, *usim\_sqn*, *usim\_rand*, *usim\_ik*, *usim\_ck*, and *usim\_res*. These fields are used to store the identity of the end-devices, the secret keys included in the USIM cards, the current state of the signaling procedure, and the different security parameters after a successful authentication.

Fig. 7 presents the log from the 3GPP Security Server upon a successful authentication, which includes the reception and transmission of the messages explained in Section V-B.

```
Starting 3GPP security server for LoRaWAN
Initialization done!
SQN for synchronization
<< Received: AttReq 00000001
>> Sending: AutReq1 5d19847cc8bde358f1a482126c97f507
<< Received: WaitAutReq2
Current SQN: 36324
>> Sending: AutReq2 de3703a8270b8000233a209cefdb869f
<< Received: AutRes 3fd7eae1d27681f0
Increasing SQN and updating NwkSKey and AppSKey with
CK ( da3ff9147cb4b9dfda3f897f991c62c7 ) and
IK ( 1f01851dd40fd7605c759aca576d60c8 )
>> Sending: AutResACK
Signaling messages      Session keys derived from 3GPP AKA procedure
```

Fig. 7. Log from the 3GPP Security Server.

Similarly, Fig. 8 presents the log from the end-device, which includes the transmission and reception of signaling messages and the communication with the USIM card. All the APDUs, for both commands and responses, are included for the sake of completeness.

## VII. CONCLUSIONS

In this work, it is proposed a low-cost solution to improve hardware security in LoRaWAN by using USIM cards as cheap cryptographic chips. This solution may be especially suitable for mobile network operators, who already have available these elements. The proposal includes the communication between the end-device and the USIM card, and the design of a security server. Both components allow to implement a 3GPP-like AKA procedure, which performs authentication and derivation of integrity and ciphering keys (used as LoRaWAN *NwkSKey* and *AppSKey* session keys). It has been employed the 4G authentication procedure, and it is left for future work to introduce 5G authentication. Since USIMs are tampered-resistant, end-devices will be no longer vulnerable to security threats such as impersonation or cloning attacks.

## ACKNOWLEDGMENTS

These last four years have been full of much effort and sacrifice and if I have managed to overcome them successfully it has been thanks to the people who have remained by my side. For that, I want to thank Jorge, my tutor, for his permanent guidance and advice throughout this project and also for having placed his trust in me. Thanks to my parents Alfonso and Josefa, my sister Lorena and my brother Alfonso for your love, concern and encouragement whenever I have needed it. I also want to thank my friends and my colleagues for all your help and support.

```

Initializing LoRaWAN (ESP32 Chip ID: 71438E2DE6B4)

>> Sending: AttReq 00000001          Signaling messages
Packet queued, waiting ACK

Received ack with 41 bytes of payload
<< Received: AutReq1 5d19847cc8bde358f1a482126c97f507
RAND: 5d19847cc8bde358f1a482126c97f507

>> Sending: WaitAutReq2 00000001
Packet queued, waiting ACK

Received ack with 41 bytes of payload
<< Received: AutReq2 de3703a8270b8000233a209cefdb869f
AUTN: de3703a8270b8000233a209cefdb869f

Performing 3G authentication...
Smartcard found - reset to trigger an ATR...
=====
SMART CARD ACTIVATION: Received ATR
=====
00|3B 9F 96 80 1F C7 80 31 A0 73 BE 21 13 67 43 20
10|07 18 00 00 01 A5
=====
SELECT EF_DIR          APDUs exchanged
                        between LoRaWAN
                        device and USIM card
=====
Status words: 6124
GET_RESPONSE
Status words: 9000
00|62 22 82 05 42 21 00 26 02 83 02 2F 00 A5 06 C0
10|01 00 CA 01 80 8A 01 05 8B 03 2F 06 04 80 02 00
20|4C 88 01 F0
READ_RECORD
Status words: 9000
00|61 19 4F 10 A0 00 00 00 87 10 02 FF FF FF FF 89
10|07 09 00 00 50 05 55 53 69 6D 31 FF FF FF FF FF
20|FF FF FF FF
=====
SELECT ADF_USIM
=====
Status words: 6159
GET_RESPONSE
Status words: 9000
00|62 57 82 02 78 21 83 02 7F FF 84 10 A0 00 00 00
10|87 10 02 FF FF FF FF 89 07 09 00 00 A5 16 83 02
20|7F FF CB 0D 00 00 00 00 00 00 00 00 00 00 00 00
30|00 CA 01 80 8A 01 05 AB 15 80 01 01 A4 06 83 01
40|0A 95 01 08 80 01 40 97 00 80 01 06 90 00 C6 09
50|90 01 40 83 01 01 83 01 81
=====
AUTHENTICATE
=====
Status words: 6135
GET_RESPONSE
Status words: 9000
00|DB 08 3F D7 EA E1 D2 76 81 F0 10 DA 3F F9 14 7C
10|B4 B9 DF DA 3F 89 7F 99 1C 62 C7 10 1F 01 85 1D
20|D4 0F D7 60 5C 75 9A CA 57 6D 60 C8 08 43 74 6F
30|BC 66 CA 6C B0

USIM authentication finished.  3GPP security params
RES: 3F D7 EA E1 D2 76 81 F0
CK: DA 3F F9 14 7C B4 B9 DF DA 3F 89 7F 99 1C 62 C7
IK: 1F 01 85 1D D4 0F D7 60 5C 75 9A CA 57 6D 60 C8

>> Sending: AutRes 00000001 3fd7eae1d27681f0
Packet queued, waiting ACK          Signaling messages

Received ack with 10 bytes of payload
<< Received: AutResACK

Changing LoRaWAN keys... security procedure finished.

```

Fig. 8. Log from the end-device.

## REFERENCES

- [1] K. L. Lueth, "State of the IoT 2018: Number of IoT devices now at 7B -market accelerating, 2018," available at <https://iot-analytics.com/state-of-the-iot-update-q1-q2-2018-number-of-iot-devices-now-7b/> (last accessed on 15/04/2019).
- [2] A. Nordrum, "Popular Internet of Things forecast of 50 billion devices by 2020 is outdated," IEEE Spectrum's General Technology Blog, 2016, available at <http://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated> (last accessed on 15/04/2019).
- [3] Arduino, available at: <https://www.arduino.cc/>
- [4] SparkFun, available at: <https://www.sparkfun.com/products/15087>
- [5] Y. Jeon, H. Ju and S. Yoon, "Design of an LPWAN communication module based on secure element for smart parking application," 2018 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, 2018.
- [6] N.Sornin, et al., "LoRaWAN Specification v1.0.2," LoRa Alliance Standard specification, July 2016, available at <https://loralliance.org/resource-hub/lorawan-specification-v102> (last accessed on 15/04/2019).
- [7] Semtech, "AN1200.22, LoRa Modulation Basics," Application Note, May 2015, available at <https://www.semtech.com/uploads/documents/an1200.22.pdf> (last accessed on 15/04/2019).
- [8] LoRa Alliance (Gemalto, Actility, and Semtech), "LoRaWAN Security, Full End-to-End Encryption for IoT Application Providers," White Paper, Feb. 2017.
- [9] 3GPP Tech. Spec. Group Service and System Aspects, 3GPP System Architecture Evolution (SAE); Security Architecture (Rel-15), 3GPP TS 33.401 V15.0.0, June 2017.
- [10] "SCLib - Smartcard access lib for Arduino based boards connected directly to a smartcard," available at <https://sourceforge.net/projects/arduinosclib/> (last accessed on 1/10/2018).
- [11] "ETSI Tech. Spec.; Smart cards; UICC-Terminal interface; Physical and logical characteristics (Release 15), ETSI TS 102.221 v15.2.0.," May 2019, available at [https://www.etsi.org/deliver/etsi\\_TS/102200\\_102299/102221/15.02.00\\_60/ts\\_102221v150200p.pdf](https://www.etsi.org/deliver/etsi_TS/102200_102299/102221/15.02.00_60/ts_102221v150200p.pdf) (last accessed on 15/10/2018).
- [12] "Getting Started with Arduino and Genuino MEGA2560," available at <https://www.arduino.cc/en/Guide/ArduinoMega2560> (last accessed on 1/10/2018).
- [13] "ESP32's guide," available at <https://www.espressif.com/en/products/hardware/esp32/overview> (last accessed on 1/10/2018).
- [14] "WiMOD Lite Gateway - Quick Start Guide v1.5," IMST GmbH, Nov. 2016, available at [https://wireless-solutions.de/downloads/Evaluation-Tools/LoRa\\_Lite\\_Gateway/WiMOD\\_LiteGateway\\_QuickStartGuide\\_V1\\_5.pdf](https://wireless-solutions.de/downloads/Evaluation-Tools/LoRa_Lite_Gateway/WiMOD_LiteGateway_QuickStartGuide_V1_5.pdf) (last accessed on 15/05/2019).
- [15] "The Things Network: iC880a-based gateway," available at <https://github.com/ttn-zh/ic880a-gateway> (last accessed on 15/05/2019).
- [16] "LoRa Server, open-source LoRaWAN network-server," available at <https://www.loraserver.io/> (last accessed on 15/05/2019).

# Sistema de bajo coste basado en bobinas para la monitorización de la humedad del suelo en entornos agrícolas

Tutora: Sandra Sendra Compte; e-mail: ssendra@ugr.es  
Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

Autor: Pablo Rodríguez Martín, e-mail: pablorodrimar@correo.ugr.es

**Resumen** — En el presente documento se expone la solución de un sensor de bajo coste basado en la inducción mutua entre bobinas para la medición de la humedad porcentual de un suelo agrícola a diferentes niveles de profundidad en tiempo real. Después de realizar un estudio del mercado actual de sensores de humedad del suelo se ha concluido en la escasa eficiencia y durabilidad de ciertos modelos mientras que otros que proporcionan un mejor resultado elevan considerablemente el coste de adquisición e implementación de una red sensórica para tal fin. Es por esto por lo que se ha llevado a cabo una investigación sobre el comportamiento de las inductancias a través de su experimentación con la orientación y relación entre ellas bajo tierra. El principal objetivo ha sido lograr una combinación óptima que permitiera obtener el nivel de humedad de la forma más lineal posible y con un umbral de tensión suficiente para su lectura por un microcontrolador ESP32, todo ello con el menor consumo de potencia posible para alcanzar una mayor autonomía. Esto se ha logrado con un modelo de dos bobinas sobre un mismo eje formadas por hilo de cobre esmaltado con un calibre de 0.4mm, enrollado en una relación 1:1 de 40 espiras cada una y operando a una frecuencia de 210kHz. A partir de ello, se ha desarrollado el esquema electrónico basado en el circuito integrado 555 para convertir la señal DC de una pila de 9V a una señal AC con la que poder inducir voltaje sobre un circuito de lectura del dato en función del medio. Adicionalmente, se ha implementado un esquema de red basado en la tecnología inalámbrica IoT de LoRaWAN que permite la recepción de dichos datos provenientes de las diferentes motas sensoras para su posterior procesamiento centralizado en una Raspberry Pi 3B+ mediante la solución *open source* de ChirpStack y su integración con InfluxDB y Grafana para su indexado, visualización y establecimiento de alarmas que notifiquen de un exceso y/o defecto en el nivel de agua del cultivo.

**Palabras clave** — bajo coste, bobinas, ChirpStack, entornos agrícolas, Grafana, humedad del suelo, inducción electromagnética, IC555, InfluxDB, Internet of Things, LoRaWAN, monitorización, mota sensora.

## I. INTRODUCCIÓN Y OBJETIVOS

CON el final de la glaciación del periodo Mesolítico que desencadena en la consumación de una vida nómada, el ser humano comienza a asentarse en torno a los primeros núcleos urbanos para poder subsistir de sus propios cultivos

en sociedad. De ello nace la necesidad histórica de los agricultores de controlar el riego a diferentes niveles del suelo, ya que el exceso o el defecto de agua pueden dañar la producción. A día de hoy, con las grandes extensiones latifundistas y las técnicas de irrigación como es el caso del riego a manta, además del gran impacto medioambiental que esto conlleva agravado por el cambio climático, también generan un elevado coste económico por la apertura de compuertas y el consumo del agua. Existen diversas metodologías y tecnologías con las que a lo largo de los años se ha podido conocer un valor aproximado de la humedad del suelo aunque, poder adaptarlas a los nuevos tiempos para controlar esto de forma remota y automatizada implica una importante inversión económica en instrumentos de medición o, en su defecto, tener que conformarse con el bajo rendimiento y duración de aquellos de menor coste.

La propuesta tecnológica que se recoge en este artículo se centra en el fenómeno de la inducción electromagnética para el desarrollo de un sensor basado en bobinas capaz de medir de una forma más eficiente la humedad del suelo con un menor coste asociado en comparación con otras soluciones más accesibles del sector IoT. Esto requiere conocer cuáles son los efectos de la humedad sobre la composición de la tierra. Dado que se pretende lograr su total autonomía, es muy importante dotar en el despliegue de estos sensores de circuitos de alimentación independientes capaces de generar una señal de corriente AC que circule a través de las bobinas. Además, también se trata de recoger los datos de los distintos nodos distribuidos por el terreno de cultivo en un único punto central. Se ha concluido que la tecnología más apropiada para esta comunicación es LoRaWAN. El objetivo de ello es poder almacenar y procesar los datos recabados a fin de poder establecer los niveles mínimos y máximos de alarma que se pueden encontrar en un tipo de suelo para así adecuarlo al rango de confort característico de la especie para poder subsistir.

## II. LA HUMEDAD DEL SUELO

En el estudio sobre la humedad en el suelo[1] se pueden diferenciar dos tipos de aguas, pues está el agua capilar, que es aquella que puede ser absorbida por la planta a través del proceso de ósmosis o que puede llegar a drenar hacia capa

más profundas de la tierra tras haber transcurrido entre 24 y 48 horas; y el agua ligada, que es aquella que queda completamente adherida a las partículas que componen el suelo y que, por tanto, no puede ser absorbida por la planta. Por tanto, existe una capacidad máxima de retención, que es lo que se conoce como Capacidad de Campo (CC); y un punto en el que solamente aparece el agua ligada, éste es el Punto de Marchitamiento (PM). La diferencia entre ambos valores es el **agua útil** que la planta puede aprovechar. Esta cuantificación del agua útil puede variar por diferentes factores como son la composición de la tierra, la permeabilidad o la porosidad. Es por ello por lo que distinguen entre suelos más arenosos o más arcillosos, como se puede apreciar en la Fig. 1.

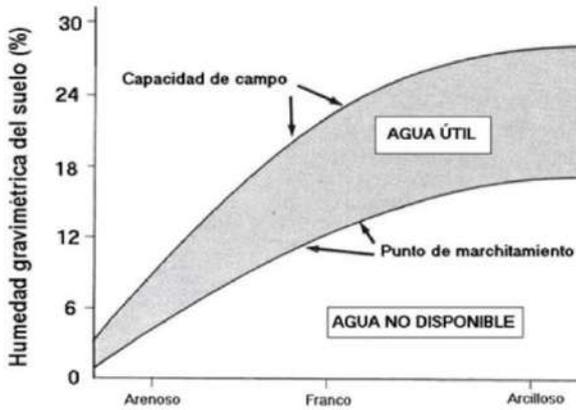


Fig. 1. Límites de agua útil por tipo de suelo

No se deben confundir estos valores con el rango de confort de la especie, pues es éste el que ha de adaptarse a las condiciones de la tierra para la supervivencia de la especie. Por tanto, en niveles porcentuales, la CC se puede definir como el 100% de humedad y el PM como el 0% de humedad. Por encima de ello estará la saturación del suelo y por debajo la más que posible asfixia radical de la planta.

### III. FUNDAMENTO TEÓRICO DEL ELECTROMAGNETISMO

A continuación se detallan los conceptos más relevantes en el análisis sobre los efectos electromagnéticos en los que se enmarca la investigación.

#### A. Principios del electromagnetismo

Según la Ley de Faraday[2] (Ec. 1), a partir de la variación del flujo magnético que atraviesa la superficie formada por un material conductor en un circuito cerrado existe de forma proporcional una *fuerza electromotriz (f.e.m.)* o voltaje inducido capaz de generar una corriente, y viceversa, formando un flujo magnético a partir de una corriente variable que circula por el conductor.

$$\varepsilon = \oint \vec{E} \cdot d\vec{l} = -\frac{d}{dt} \int_S \vec{B} \cdot d\vec{S} \quad (V) \quad (1)$$

Tal y como enunció Lenz a partir del trabajo de Faraday, esta corriente debe ser variable con respecto al tiempo ya que existe un campo magnético que se opone a este cambio por el principio de conservación de la energía[3] por lo que, de

hacerse constante, lo anula para evitar que la corriente siga aumentando o disminuyendo. Lo mismo ocurre para el campo magnético. Por tanto, en el caso de dos bobinas cercanas, una bobina primaria a través de la que circula una corriente variable en el tiempo puede inducir una corriente sobre una bobina secundaria. Este efecto de voltaje inducido se potencia con un mayor número de espiras  $N$ , según se refleja en la Ec. 2.

$$\varepsilon = \oint \vec{E} \cdot d\vec{l} = -N \frac{d}{dt} \int_S \vec{B} \cdot d\vec{S} = -N \cdot \frac{d\Phi}{dt} \quad (V) \quad (2)$$

#### B. Comportamiento de las bobinas. Inductancia

La inductancia[4] se define por tanto como la propiedad electromagnética de las bobinas de oponerse a cualquier variación de la corriente que circula a través de ellas. Cuanto mayor sea el número de espiras  $N$  y el radio  $r$  de éstas, mayor será su valor (Ec. 3). De esta forma, se almacenará la energía suministrada en forma de campo magnético cuando exista un trabajo contra el crecimiento de la corriente en la bobina y, por tanto, se invierta la polaridad de ésta, lo que se conoce como *fuerza contraelectromotriz (f.c.e.m.)*; y esta misma se irá descargando a modo de contribución energética a medida que la corriente disminuya, es decir, evitará cambios abruptos de la corriente al intentar oponerse a dicha variación.

$$L = \frac{\mu_0 \mu_r N^2 \pi r^2}{l} \quad (H) \quad (3)$$

Cuanto mayor sea la inductancia, mayor se hará la *f.c.e.m.*, aunque a medida que el campo magnético alcance su capacidad máxima, esta fuerza se irá haciendo cada vez más pequeña hasta hacerse nula permitiendo así en algunos casos circular la máxima corriente al tratarse la bobina de un corto. Esto implica que, ante la entrada de un pulso cuadrado de corriente a través de una bobina, le llevará más tiempo alcanzar el pico de corriente cuanto mayor sea la inductancia. Para el caso de señales periódicas, el hecho de aumentar la frecuencia reducirá la variación de corriente en el circuito, es decir, aumentará la reactancia inductiva (Ec. 4).

$$X_L = \omega L = 2\pi f L \quad (\Omega) \quad (4)$$

#### C. Inductores e inductancia mutua

La agrupación de inductancias es muy similar a la de los resistores, siendo la inductancia total la suma de todos sus valores cuando están en serie y la inversa de la suma de sus admitancias cuando están en paralelo.

El esquema de una bobina real se suele representar como esta inductancia en serie con la impedancia característica del hilo conductor y ambas en paralelo con una capacitancia parásita debida a la fina capa de esmalte que se encuentra entre las espiras, como se muestra en la Fig. 2.

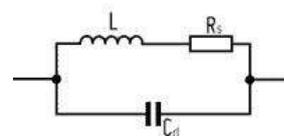


Fig. 2. Modelo equivalente de una bobina real

Por esta capacidad, se presentará un pico de impedancia entre los extremos de la bobina en la que se conoce como frecuencia de resonancia,  $f_r$ , donde la corriente de realimentación que carga y descarga tanto a bobina como a condensador en este circuito RLC establece un equilibrio de atenuación y fase con respecto a la suministradora. Esta  $f_r$  puede calcularse mediante la Ec. 5.

$$f_r = \frac{1}{2\pi} \sqrt{\frac{1}{LC_d} - \frac{R_s^2}{L^2}} \approx \frac{1}{2\pi\sqrt{LC_d}} \text{ (Hz)} \quad (5)$$

De esta forma, si se representa la impedancia con respecto a la frecuencia, se obtiene un resultado similar al de la Fig. 3. En ella se puede ver cómo aparece un comportamiento inductivo para frecuencias más bajas que la de resonancia mientras que para aquellas más altas se presenta un comportamiento capacitivo.

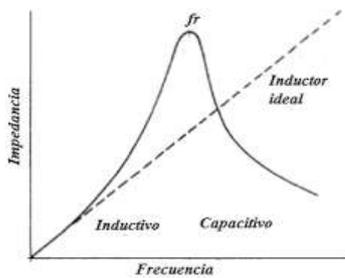


Fig. 3. Comportamiento real de una bobina

Para la  $f_r$ , el valor de la impedancia se eleva sobre el caso de un inductor ideal, aumentando así el voltaje inducido. Este valor dependerá de lo que se conoce como *factor de calidad*, el cual se calcula según la Ec. 6.

$$Q = \frac{X_L}{R} = \frac{2\pi f L}{R} \quad (6)$$

Sin embargo, esto no implica unilateralmente a cada bobina, sino que se produce una interdependencia entre ambas, la primaria y la secundaria, lo que se conoce como inductancia mutua  $M$ . Ésta puede calcularse como en la Ec. 7.

$$M = k\sqrt{L_1 L_2} \text{ (H)} \quad (7)$$

Este factor de acoplamiento  $k$  es el que hace que esa frecuencia de resonancia se desplace hacia frecuencias inferiores y superiores, comenzando antes ese pico de tensión máxima inducida, tal y como se puede apreciar en la Fig. 4.

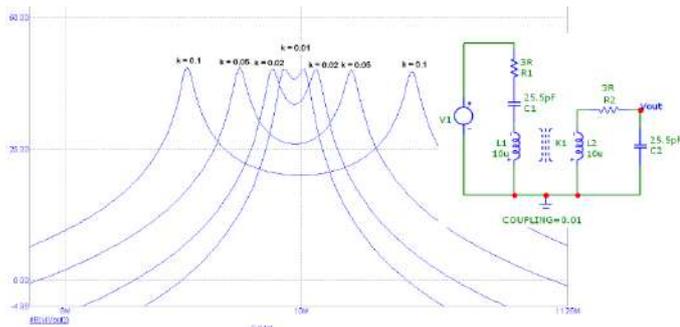


Fig. 4. Resonancia con inducción mutua  $M$  en función de  $k$

#### D. Inducción mutua en función del medio

Además, como se ha visto anteriormente (Ec. 3), la inducción entre bobinas se producirá a través de un medio material, principalmente caracterizado por la permeabilidad magnética relativa ( $\mu_r$ ) [5] del núcleo que exista dentro de ellas. Estos materiales se clasifican en los de la Fig. 5.

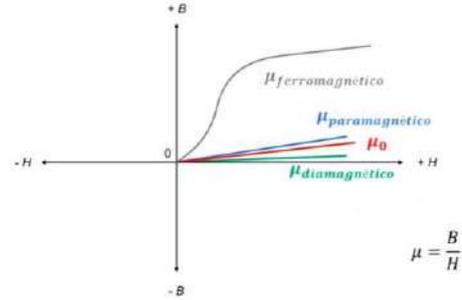


Fig. 5. Permeabilidad magnética en función del medio material

- **Ferromagnéticos** ( $\mu_r \gg 1$ ). Absorben parte del campo magnético y se convierten en fuentes canalizadoras del mismo, intensificando el campo magnético y las corrientes inducidas.
- **Paramagnéticos** ( $\mu_r > 1$ ). Favorecen en muy poca medida este comportamiento.
- **Diamagnéticos** ( $\mu_r < 1$ ). Repelen en muy poca medida los campos magnéticos.

De este modo, la presencia de agua en la tierra alterará el medio en el que se ven rodeadas las bobinas en función de la composición de este líquido, por lo que se pretende cuantificar este cambio durante la experimentación.

#### IV. REALIZACIÓN PRÁCTICA

La fase experimental de la investigación se ha dividido en el diseño del circuito de alimentación; la propia experimentación con diferentes niveles de humedad y frecuencias para diferentes prototipos de bobinas; y la implementación de la red LoRaWAN para los datos.

##### A. Diseño del circuito de alimentación

Dado que el principal objetivo del circuito de alimentación consiste en generar sobre las bobinas una señal de corriente AC a partir de una pila de 9V, se ha optado por el uso del circuito integrado 555 en configuración *Astable*. De este modo, no hay ninguna señal externa que intervenga en la creación de pulsos sino que, modificando los valores de las resistencias  $R_A$ ,  $R_B$  y la capacidad  $C$  de la Fig. 6, se puede establecer un oscilador capaz de operar con diferentes frecuencias para un *duty cycle* entre el 50% y el 100%.

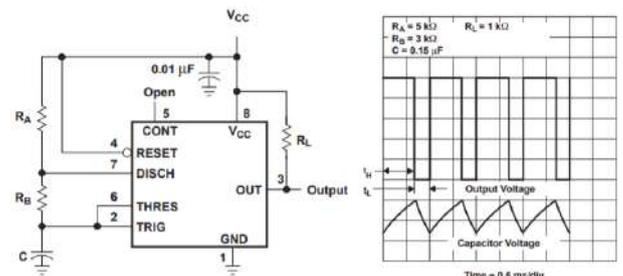


Fig. 6. Modo de operación Astable del IC555

Esta frecuencia y el *duty cycle* vienen dados por las expresiones de la Ec. 8 y la Ec. 9, respectivamente.

$$f \approx \frac{1.44}{(R_A+R_B)C} \text{ (Hz)} \tag{8}$$

$$D = \frac{R_A+R_B}{R_A+2R_B} \cdot 100 \text{ (\%)} \tag{9}$$

Sin embargo, este circuito integrado posee limitaciones en frecuencia. En el caso del NE555P de Texas Instruments[6], esta frecuencia límite está en torno a los 100kHz, aunque otros modelos basados en la tecnología de transistores CMOS permiten lograr mayores frecuencia, como es el caso de los 500kHz del circuito integrado ICM7555 adquirido. Por tanto, conviene encontrar los picos de resonancia en bobinas dentro de ese rango. Con todo ello, el circuito de alimentación final es el que se puede ver en la Fig. 7.

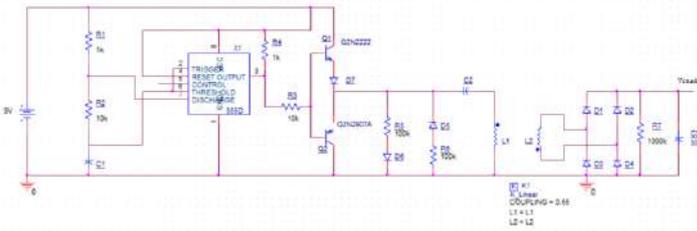


Fig. 7. Esquemático del circuito de alimentación inversor

Para lograr invertir la señal DC a una AC quasi-sinusoidal, se recurre a la sincronización de la resonancia entre la bobina primaria  $L_1$  y la capacidad  $C_2$  con la conducción y corte de los transistores BJT. De este modo, para un valor alto de la señal de pulsos a la salida del ICM7555, por el transistor NPN comenzará a circular una corriente a causa de la *f.c.e.m.* de la bobina que se irá incrementando a medida que ésta se vaya cargando e irá decrementándose cuando el condensador comience a alcanzar su máxima carga. Del mismo modo, para un valor bajo, el transistor NPN pasará a estar en corte mientras que el PNP comenzará a conducir, circulando la misma corriente de carga en sentido inverso por la rama hasta masa. De esta forma se puede lograr una corriente AC sobre la bobina primaria para que pueda conducir una corriente proporcional sobre la bobina secundaria, tal y como se ha visto en los fundamentos teóricos de electromagnetismo. En la Fig. 8 se puede ver la corriente que circula a través de los transistores (azul y morado), como la que circula a través de la bobina (magenta).

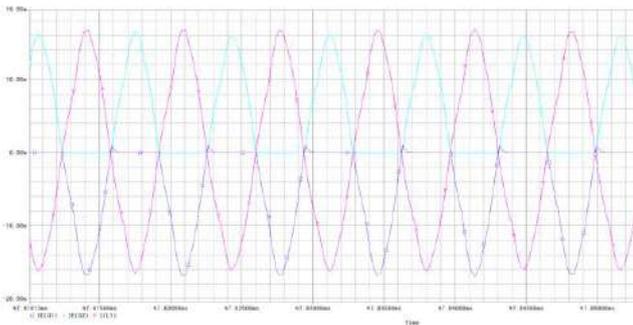


Fig. 8. Gráfica de corrientes en transistores y en la bobina primaria para una frecuencia de 160kHz (50%),  $L_1$  de 1mH y  $C_2$  de 990pF

Adicionalmente, se ha provisto al circuito de dos ramas de alta impedancia para evitar dañar la circuitería de no estar en total resonancia, pues habrá que adaptar esa capacidad  $C_2$  a la bobina primaria utilizada, por lo que la señal de corriente podrá diferir a la anterior. La tensión de salida DC para la lectura del microcontrolador ESP32 establecida mediante un rectificador de diodos se prevé que se establezca transcurrido un corto periodo de tiempo, como se aprecia en la Fig. 9.

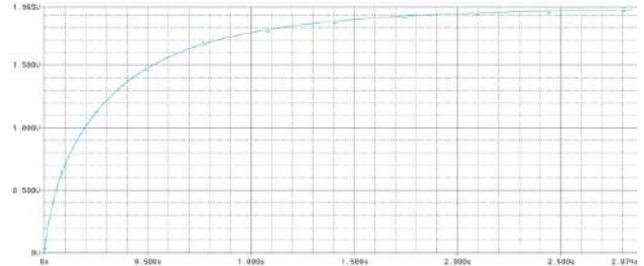


Fig. 9. Valor de tensión de salida del sistema en función del tiempo

**B. Experimentación con bobinas**

En la experimentación se han desarrollado cuatro prototipos de bobina mediante el arrollamiento de hilo de cobre esmaltado de 0.4mm sobre tubos huecos de material PVC. Los prototipos son los detallados en la Tabla I.

TABLA I. PROTOTIPOS DE BOBINAS PARA EL SISTEMA SENSOR

Prototipo	Capas	N1	N2	Relación	Diámetro
P1	1	40	40	1:1	28mm
P2	1	40	80	1:2	28mm
P3	1	40	100	1:2.5	28mm
P4	4	40	80	1:2	43mm

El procedimiento llevado a cabo consiste en la medición del voltaje de salida del sistema de la Fig. 7 tras sustituir las bobinas por las de los diferentes prototipos en un suelo franco y realizarlo para un barrido de frecuencias en cada nivel de humedad porcentual. Para establecer estos niveles de humedad (0%, 25%, 50%, 75% y 100%) se ha recurrido al método gravimétrico de la Ec. 10.

$$\text{Humedad (\%)} = \frac{m_{\text{húmedo}} - m_{\text{seco}}}{m_{\text{seco}}} \cdot 100 \tag{10}$$

Por ello, para unos 300g de tierra, se añaden 75ml de agua para aumentar el nivel, hasta llegar a un peso total del sistema de aproximadamente 600g. Es importante mencionar que es complicado llegar al 100% de la humedad ya que, como se ha comprobado en la Fig. 1, la CC de cualquier tipo de tierra está muy por debajo de dicho valor. En la Fig. 10 puede verse el sistema utilizado durante la experimentación.



Fig. 10. Prototipos reales de bobinas utilizados para la experimentación

Dada la situación de pandemia ocasionada por el virus SARS-CoV-2 y el decreto del Estado de Alarma en España, ha sido imposible acceder al laboratorio de la ETSIIT para llevar a cabo las mediciones correspondientes, por lo que esta investigación se ha tenido que adaptar a las circunstancias sustituyendo el instrumental de laboratorio como puede ser el generador de señales o el osciloscopio por una versión reajustada del circuito de alimentación y un multímetro capaz de medir también la frecuencia de la señal para establecer una relación con el voltaje inducido. Este nuevo circuito puede apreciarse en la Fig. 11, para el que se han sustituido los transistores de inversión y la rama resonante por un único transistor NPN conectado a la bobina con diodos a modo de *freewheeling*. De esta forma, sustituyendo la capacidad principal del circuito mediante un surtido de condensadores en serie y paralelo se pueden lograr diversas frecuencias. Por ende, pese a haber simplificado el circuito, ya no se logra una corriente AC en la bobina, pero sí es variable, por lo que se puede proceder al estudio de la inducción entre bobinas.

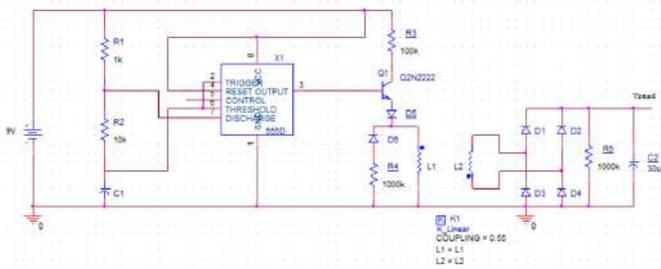


Fig. 11. Esquemático del circuito de experimentación con bobinas

El mejor resultado obtenido para los diferentes casos ha sido el del primer prototipo, con una relación 1:1 de 40 espiras, debido a la mayor linealidad obtenida para una frecuencia de **210kHz**, tal y como se puede concluir con las gráficas de la Fig. 12. En ellas se representa el valor de tensión de salida frente a las diferentes frecuencias logradas y a los diferentes niveles de humedad.

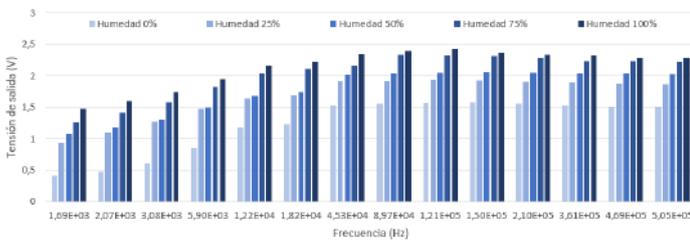


Fig. 12(a). Experimentación con el Prototipo 1

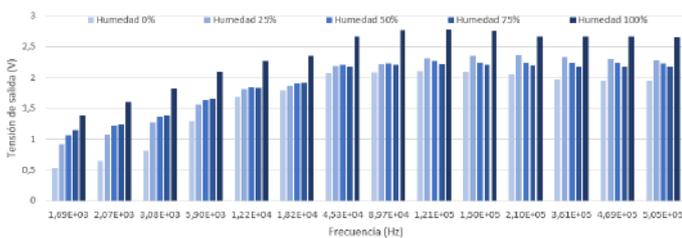


Fig. 12(b). Experimentación con el Prototipo 2

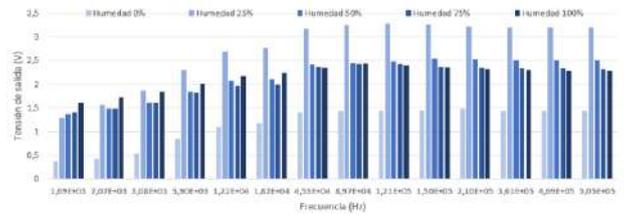


Fig. 12(c). Experimentación con el Prototipo 3

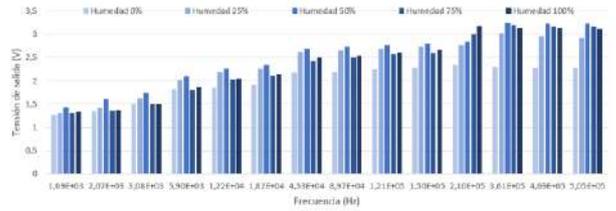


Fig. 12(d). Experimentación con el Prototipo 4

La no-linealidad del resto de prototipos, pese a lograr un mayor valor de tensión inducida frente al primer prototipo, los hace no aptos para el sensor a implementar. Cuanto más se aproxime esta máxima linealidad al pico de resonancia, más se logrará optimizar la potencia de consumo del sensor. Nótese que el hecho de tener un crecimiento o un decrecimiento a medida que se eleve la humedad no es relevante ya que solamente interesa conocer en qué grado afecta la presencia de agua en la tierra donde se encuentra el sensor y no de la propia composición de agua y del suelo. Esta experimentación se ha de realizar por tanto para cada tipo de suelo, particularizando cada caso. La temperatura también afecta tal y como se describe en el libro *Interferencias electromagnéticas en sistemas electrónicos*[7].

Una vez seleccionado qué prototipo se utilizará en este proyecto, conviene integrarlo con el circuito original de alimentación de la Fig. 7. Ello requiere adaptar también la capacidad C2 para lograr el mayor voltaje de salida posible, acercándose así a la resonancia comentada. Para un valor de **20.4nF** se ha logrado un valor máximo de tensión de 2.98V. Sin embargo, experimentando con las limitaciones de los pines ADC del ESP32 adquirido se ha visto que éste no responde de forma lineal a partir de los 2.8V, por lo que se ha de reemplazar la resistencia R7 del circuito por una de menor valor para así limitar dicha tensión de salida por debajo de este valor. En este caso se ha reemplazado por una resistencia de **10kΩ**. Si ahora vuelve a realizarse las mediciones para los diferentes niveles de humedad se obtiene la siguiente gráfica de la Fig. 13.

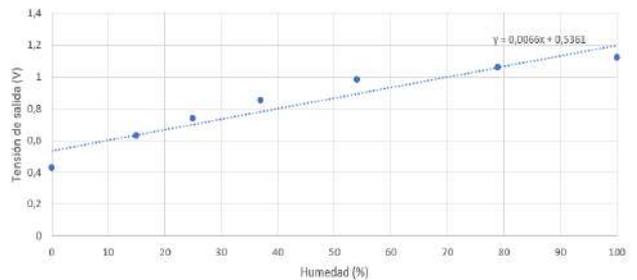


Fig. 13. Valores de tensión de salida del sensor para la humedad del suelo

Como se puede ver, se logra un crecimiento aproximadamente lineal hasta el 54% de la humedad, valor a partir del cual este sistema deja de crecer con dicha pendiente para establecerse en torno a la unidad del voltio entre el 79% y el 100%. Como se ha visto antes, ningún suelo es capaz de retener un 100% de humedad, pues el agua dreña hacia más allá de la profundidad del sensor y se escapa del recipiente, con lo que este sensor se aproxima bastante bien a la realidad. En este caso, con una tierra más bien franca, basta con que el sistema sea preciso hasta el 25-30%, pues por encima de ello el suelo se saturará y el agua drenará en cuestión de horas.

### C. Implementación de la red LoRaWAN

Para probar la red LoRaWAN que llevará los datos recolectados de los diferentes nodos hasta un punto central situado a media distancia se han utilizado dos módulos ESP32-LoRa, uno de ellos como nodo final programado para la lectura y transmisión del dato; y el otro como *gateway* monocanal para su recepción. La programación de ambos nodos puede encontrarse en el repositorio de GitHub[8]. Esta red se gestiona a través de los servidores de red y de aplicación de ChirpStack instalados en una Raspberry Pi 3B+, conectada al *gateway* a través del protocolo UDP de forma inalámbrica con la tecnología WiFi. Los nodos finales se han configurado en clase A para lograr un menor consumo. Por último, se ha integrado este sistema con la base de datos de *InfluxDB* para indexar los datos recogidos con sello temporal y se ha configurado la aplicación de *Grafana* para su visualización y el establecimiento de alarmas en el caso de que en nivel de humedad salga de cierto intervalo (eg. rangos de confort de la planta), como se ha hecho en este caso para un 20%.

### V. PRUEBA FINAL DEL PROYECTO

Conectando la mota sensora independiente tal y como aparece en la Fig. 14 e inicializando tanto el *gateway* como los servidores alojados en la Raspberry Pi 3B+, ya se puede pasar a probar el funcionamiento general del esquema.

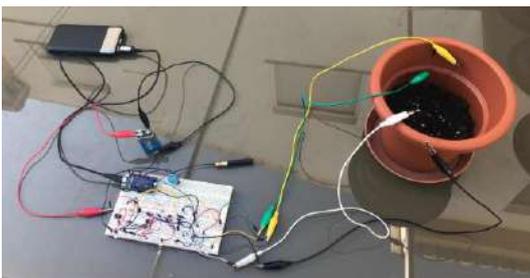


Fig. 14. Prueba del sistema de monitorización de humedad del suelo

Partiendo de tierra totalmente seca, se ha probado a humedecerla transcurrido un tiempo tras inicializar el servicio de monitorización construido. Primero se ha vertido una pequeña cantidad de agua para asegurarse que el sensor era ligeramente excitado. Luego se ha llenado hasta cubrir con agua la totalidad de la superficie de la tierra, lo que equivale a un 100% de la humedad. El resultado ha sido el de la Fig. 15. Nótese que, antes de haber vertido el agua, se ha lanzado una alerta tras tomar dos muestras consecutivas en las que se ha detectado un valor de humedad inferior al 20%.



Fig. 15. Datos recopilados en la aplicación

### VI. CONCLUSIONES Y VÍAS FUTURAS DE ESTUDIO

En este proyecto se ha logrado obtener un prototipo de sensor de humedad del suelo basado en los principios electromagnéticos de las bobinas así como del circuito de alimentación asociado a la mota sensora; y la red de datos que permite monitorizar en todo momento este parámetro en diferentes puntos del cultivo de forma remota, segura y en tiempo real. Además, se trata de una solución de muy bajo coste, aunque es necesaria la repetición de todo este proceso de investigación para cualquier lugar donde se desee implantar este sistema.

Como líneas futuras de trabajo se encuentra el estudio de otros nuevos prototipos de bobinas; la experimentación con otros tipos de tierras más arenosas y arcillosas para encontrar posibles patrones que permitan generalizar su uso; y estudiar el grado de afección entre bobinas para optimizar su distribución a diferentes profundidades del suelo.

### AGRADECIMIENTOS

Quiero mostrar mi especial agradecimiento a mi familia y a amigos/as, así como a todos/as los/as compañeros/as y profesores/as por todo el apoyo recibido.

### REFERENCIAS

- [1] Dr. Juan José Ibáñez. Niveles de humedad del suelo y agua disponible: El agua en el suelo 3. Madrid+D Blogs, junio 2006. <https://www.madrimasd.org/blogs/universo/2006/06/26/33002>, último acceso: 06/07/20.
- [2] Ricardo Trumper & Moshe Gelbman. Investigating electromagnetic induction through a microcomputer-based laboratory. *Physics Education*, 32(2):90-95, febrero 2000.
- [3] Greg Gbur, Daniel James & Emil Wolf. Energy conservation law for randomly fluctuating electromagnetic fields. *Physical Review E*, 59(4):4594, 1999.
- [4] F. B. J. Leferink. Inductance calculations; methods and equations. Pages 16-22, agosto 1995.
- [5] Charles Kittel. Theory of the dispersion of magnetic permeability in ferromagnetic materials at microwave frequencies. *Phys. Rev.*, 70:281-290, septiembre 1946.
- [6] 555 IC datasheet. *Texas Instruments*. <http://www.ti.com/lit/ds/symlink/ne555.pdf>, último acceso: 06/07/20.
- [7] Josep Balcells. Interferencias electromagnéticas en sistemas electrónicos. Marcombo, 2005.
- [8] Pablo Rodríguez Martín. Repositorio GitHub del proyecto TFGcode. <https://github.com/paroma96/TFGcode>, último acceso: 06/07/20.



**Pablo Rodríguez Martín.** Nacido en Granada el 18 de abril de 1996. Graduado en Ingeniería de Tecnologías de Telecomunicación con mención en Telemática por la Universidad de Granada. Actualmente estudiante del Máster Oficial Universitario de Ingeniería de Telecomunicación de la Universidad de Granada. Realizó una estancia Erasmus+ durante el curso 2017-2018 en la Politechnika Warszawska (Polonia).

# Sistema de control para UAVs y reconocimiento de objetos

Autor: Joaquín Lara Sarabia, e-mail: joaquinlara@correo.ugr.es  
Tutor: Pablo Ameigeiras Gutiérrez, e-mail: pameigeiras@ugr.es  
Cotutor: Juan José Ramos Muñoz, e-mail: jjramos@ugr.es  
Titulación: Ingeniería de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—En esta memoria se describe el diseño, montaje y programación de un vehículo aéreo no tripulado (UAV) y sus sistemas de control, las cuales son tareas complementarias que requieren conocimientos y competencias de distintos ámbitos, varios de ellos relacionados con las Tecnologías de la Información y las Comunicaciones.

En este proyecto se diseñado, ensamblado y programado un UAV del tipo cuadricóptero a partir de componentes fáciles y baratos de conseguir, como son placas de arduino y sensores/actuadores compatibles con este microcontrolador. El cuadricóptero es controlado por WiFi, y dispone de una cámara y un módulo de localización GPS.

Asimismo, se ha desarrollado un servidor y una aplicación Android que permiten a varios o a un mismo usuario gestionar distintos UAV. El servidor, además, realiza reconocimiento de objetos de las imágenes capturadas por el UAV. Por otra parte, el usuario puede controlar su(s) UAV mediante el acelerómetro y giroscopio del dispositivo Android. La aplicación Android, el servidor y el propio UAV se comunican mediante un protocolo diseñado expresamente para este sistema.

**Palabras clave**—Android, Arduino, Cuadricóptero, Construcción, Dron, Protocolo de Comunicación, Reconocimiento de Objetos, Servidor de Gestión, Simulador, Sistema de Control, UAV, WiFi

## I. INTRODUCCIÓN

UN dron, conocido formalmente como vehículo aéreo no tripulado (VANT), o UAV en inglés (*Unmanned Aerial Vehicle*) es una aeronave que vuela sin tripulación, ya sea controlada de forma remota por un operador o bien por control automático. Este tipo de vehículos son cada vez más populares y su importancia está creciendo en sectores tan relevantes como en la industria o el ámbito militar. Asimismo, gozan de una gran cantidad de aplicaciones, algunas de ellas muy conocidas en la sociedad, como son la fotografía y el vídeo. Se trata de una tecnología sin limitaciones, pues otro tipo de aplicaciones podría ser la ayuda en el control de catástrofes, aplicaciones en el campo de la medicina o el medio ambiente o incluso como ayuda a personas con discapacidad.

Este TFG se ha centrado en el diseño, montaje y programación de un cuadricóptero, que es un tipo especial de UAV multi-rotor y con cuatro brazos, donde en el extremo de cada uno de ellos se encuentra un motor que hace girar una hélice, por comodidad en la redacción y la lectura, se ha denominado dron a este cuadricóptero. Para el diseño y la construcción del mismo se ha empleado un microcontrolador

*open-hardware* cuyo uso está muy extendido: Arduino, el cual presenta un gran número de limitaciones que han sido necesarias resolver para poder llevar a cabo el proyecto.

Además del cuadricóptero, también ha sido necesario diseñar y construir el sistema de comunicación completo, formado por un servidor de gestión y un protocolo de comunicación. Este servidor de gestión tiene integrado además un servidor de reconocimiento de objetos, que permite capturar y analizar las capturas realizadas por el dron, y un servidor de simulación, que permite trabajar y manipular drones simulados.

Por último, el sistema de control está compuesto por una aplicación Android, que simula los movimientos del dron de forma natural mediante el uso del acelerómetro y giroscopio integrados en el propio dispositivo y envía las órdenes de control al dron para que repita los movimientos en tiempo real.

## II. ESTADO DEL ARTE

SON muchos los proyectos que pretenden el diseño y construcción de un dron utilizando para ello elementos *open-hardware* y *open-source*. Sin embargo, estos sistemas adolecían de una serie de desventajas que no hacían posible su implementación en este proyecto. Uno de los objetivos principales era ser capaz de construir un dron barato y con todas las funcionalidades posibles que ofrecen los drones de gama alta. Requisito que no cumplían los sistemas existentes, como pueden ser los *autopilots*. Dentro de estos proyectos uno de los más conocidos es *Ardupilot* [1].

Además de proyectos de *open-source*, existen una gran cantidad de drones comerciales totalmente montados y dispuestos para comenzar a volarlos, pero de nuevo se trata de una alternativa costosa y poco flexible, pues normalmente no permiten ser programados, y si es posible, son aún más caros.

Por último, mencionar que existen una gran cantidad de herramientas y simuladores desarrollados para permitir y facilitar la construcción y diseño de un dron que haga uso del protocolo *MAVLink* [2]. En este caso se ha implantado uno de estos simuladores en el servidor de gestión gracias a una traducción realizada entre el protocolo de comunicación

desarrollado en este proyecto y el protocolo *MAVLink*, que tampoco ha sido posible utilizar en este proyecto debido a su complejidad, pues las piezas utilizadas en la construcción del dron, al ser baratas, presentan una gran cantidad de limitaciones que han tenido que ser solucionadas mediante sistemas y protocolos propios.

#### A. Conclusiones del estado del arte

Existe una amplia variedad de drones, simuladores y protocolos disponibles en el mercado.

En cuanto a los drones, algunos tienen objetivos orientados a un uso más profesional sobre todo en el ámbito de la fotografía, pero tienen un precio muy elevado, mientras que otros drones más dedicados a un uso personal tienen peores características y a veces con un precio también demasiado alto para lo que ofrecen. Del mismo modo la mayor parte de todas estas opciones son propietarias, por lo que no sería posible para un desarrollador modificar ni configurar estos drones para usos más específicos.

En cuanto a los protocolos de comunicación, cada fabricante utiliza sus propios mensajes y arquitecturas de control, por lo que no son una opción viable para el desarrollo de drones propios. Está la excepción de *MAVLink*, que aunque reúne los requisitos fundamentales para el control de un dron, en casos de aplicaciones concretas interesaría un sistema de comunicación más directo y que responda de manera más eficaz a la problemática en concreto.

Por todo ello, se puede observar como no existe una solución *open-source* eficaz para resolver los objetivos planteados en este proyecto. Es decir, el desarrollo de un dron con el menor presupuesto posible pero que permita una gran flexibilidad en su fabricación, utilizando componentes variados y únicamente hardware y software libre, que permitan a cada persona ajustar el dron a sus necesidades y configuraciones específicas.

### III. DISEÑO Y MONTAJE DEL DRON

SE trata de la sección con mayor peso e importancia en el proyecto, pues en este apartado se describe todo lo necesario sobre el funcionamiento de un dron y qué es necesario para diseñar y construirlo desde cero.

#### A. Estructura y funcionamiento de un dron

En este capítulo se detallan todas las partes que componen un dron, de modo que se mostrarán además los materiales escogidos y el por qué de su elección. Se detalla también el diseño de la electrónica y el funcionamiento de los distintos componentes. En los últimos apartados se muestra todo el proceso de construcción y montaje y finalmente se detalla el software desarrollado para las placas de control integradas en el dron. Como inconveniente, no es posible mostrar todo el proceso en este resumen debido a su longitud.

Lo primero que hay que saber es de qué partes está compuesto un dron. La Figura 2 muestra dichas partes sobre

el dron construido finalmente, para tener una mejor visión de cómo es cada una de estas partes, y cómo se integran sobre la estructura global.

Se puede distinguir principalmente una base o chasis fabricado en madera y diseñado en este mismo proyecto. Como ejemplo, la Figura 1 muestra los planos utilizados para construir la base superior de la cabina del dron. Aunque existen materiales mucho más preparados e idóneos para la fabricación del dron, la madera es un material muy versátil y barato, que fue el principal motivo de su elección, ya que otros materiales como la fibra de carbono o el aluminio son muy caros y difíciles de manipular. Sobre este chasis se montan el resto de materiales, como pueden ser los motores, las hélices o las baterías. En la cabina principal se encuentran los módulos Arduino que componen el cerebro o placa controladora de vuelo. En total hay tres placas Arduino Mega2560 [3]. El hecho de que haya tres viene dado por la complejidad de las operaciones y el código a ejecutar, así como de las limitaciones de tiempo impuestas por las operaciones matemáticas necesarias para poder realizar el vuelo, como son la estabilización y controlador PID (*Proportional-Integral-Derivative*). Otro de los elementos importantes dentro del dron son los *Electronic Speed Controller* (ESC), que están conectados a los motores, la(s) batería(s) y la placa controladora de vuelo. Son los encargados de convertir los pulsos procedentes de la placa Arduino en potencia para los motores. De este modo gracias a estos dispositivos es posible controlar la potencia que llega a cada motor y por lo tanto es posible controlar la altura, dirección y estabilización del dron.

Como se ha comentado, la placa controladora de vuelo estará formada por tres placas Arduino Mega2560. Esta placa Arduino se ha escogido debido a sus altas prestaciones y al gran número de buses de comunicación serie que presentan (4), lo que permite que todas las placas se comuniquen entre ellas a la vez que ejecutan tareas paralelas convirtiéndose así en un verdadero cerebro para el dron.

Para la comunicación con el servidor de gestión y el controlador de vuelo se ha utilizado el módulo *WiFi CC3000* de *Texas Instruments* [4]. La cual, aunque cumple su función, presenta unas grandes limitaciones, como son una alta variabilidad ante el voltaje (necesita una buena fuente de alimentación y estable), unos tiempos de lectura y escritura

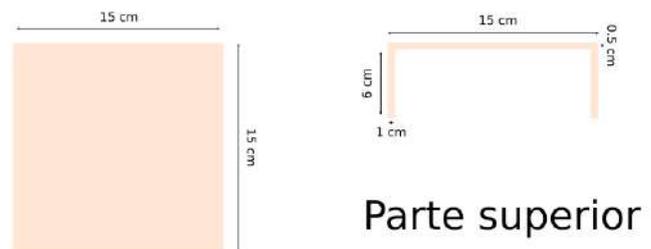


Fig. 1. Planos de la base superior para la construcción de la cabina.

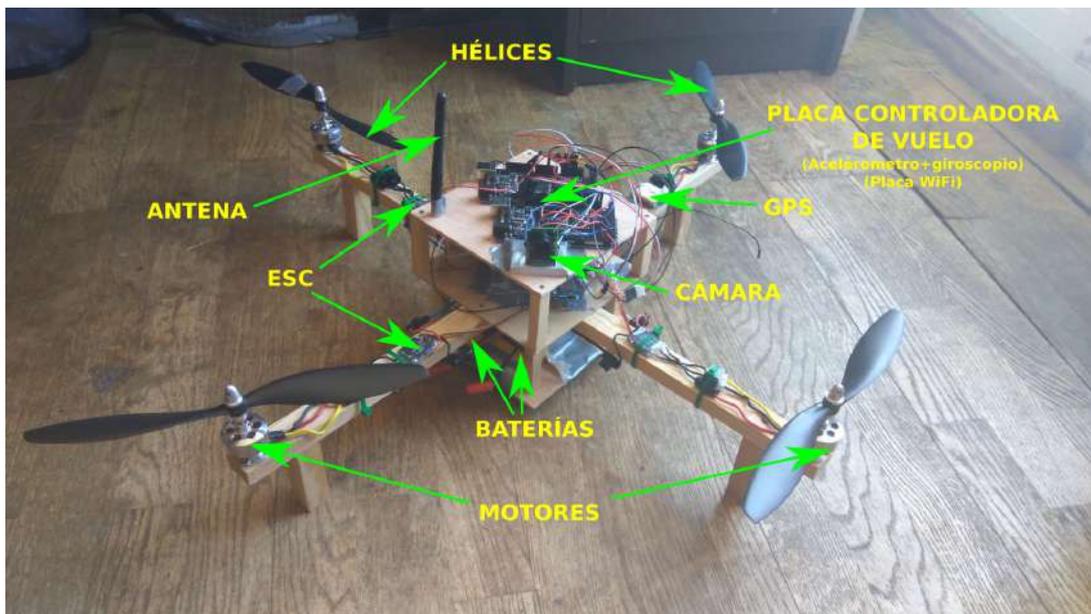


Fig. 2. Distintos componentes y partes que conforman el dron.

en los sockets muy lentos y una fuerte limitación en el número de bytes que se pueden transmitir por paquete (alrededor de 100 bytes), lo que hace que por ejemplo la transmisión de las imágenes deba ser fragmentada en un gran número de paquetes para poder ser enviada al servidor de reconocimiento de objetos.

Por último, otros de los componentes fundamentales en cualquier dron son los sensores utilizados para la estabilización del vuelo. Los drones profesionales utilizan una serie de acelerómetros y giroscopios que permiten calcular con gran precisión las aceleraciones, a partir de las cuales se estiman las desestabilizaciones y rotaciones en cualquiera de los tres ejes tridimensionales. En ocasiones, el acelerómetro y giroscopio vienen acompañados de un magnetómetro, que actúa como brújula para conocer en todo momento la dirección en la que apunta el dron. En este proyecto se ha utilizado el módulo *MPU6050* [5]. Se trata de un dispositivo de bajo consumo de potencia y de bajo coste pero que presenta un alto rendimiento. Integra tanto un acelerómetro de 3 ejes como un giroscopio. Es un sensor muy preciso, con una conversión analógica-digital de 16 bits para cada canal, lo que permite capturar los tres ejes ( $x, y, z$ ) al mismo tiempo. Es capaz de medir las aceleraciones lineales y angulares, gracias a las cuales se obtienen los ángulos de inclinación respecto a los tres ejes principales (*pitch* o eje vertical, *roll* o eje horizontal y *yaw* o eje perpendicular).

### B. Diseño del esquemático

Una vez conocidas todas las partes que conforman el dron y escogidos los componentes finales que se utilizan es necesario diseñar todo el esquema de conexiones que se va a realizar en función de la disposición de los elementos y como todos ellos se comunican entre sí. En la Figura 3 se muestra el esquemático final diseñado. En él se pueden apreciar las tres placas Arduino que conforman la placa controladora de

vuelo y cómo los distintos elementos se conectan a ellas.

Según estos elementos se puede comprobar la función que realiza cada una de estas placas:

- **Placa de comunicación:** esta placa tiene conectado el módulo WiFi *CC3000*, de modo que se encargará de enviar las peticiones e imágenes tomadas por la cámara al servidor de gestión y sobre todo más importante aún, se encargará de recibir las órdenes de control procedentes de la aplicación Android que el operario del dron envíe. La necesidad de que esta placa se encargue únicamente de las comunicaciones viene dada por las limitaciones antes descritas, donde al ser la placa tan lenta es necesario que esta placa únicamente se encargue de esta función y no del control del dron también, pues en ese caso las restricciones temporales del cálculo y mediciones para llevar a cabo la estabilización son tan elevadas que no podrían ser resueltas a tiempo. Además del módulo WiFi también se encarga de la localización mediante el módulo GPS.
- **Placa de control:** esta placa es la esencial de este proyecto, ejecuta el código necesario para la estabilización del dron. Es fundamental que el tiempo de ejecución en bucle del código de esta placa se realice en un tiempo no mayor de 3ms, esto es, una frecuencia de 333Hz, pues en caso contrario el dron se volverá poco responsivo y no será posible su control. Es este el motivo por el cual aparecen el resto de placas, ya que las operaciones que estas realizan (enviar información al servidor, recibir información, tomar capturas, recoger los datos del gps, etc) son operaciones bloqueantes que requieren de un gran tiempo de procesamiento, lo que haría que no fuese posible obtener el tiempo de ejecución de 333Hz, además de suponer problemas de alimentación al haber tantos componentes conectados. Del mismo modo esta opción hace al proyecto más flexible y escalable.
- **Placa de captura de imágenes:** esta placa tiene la

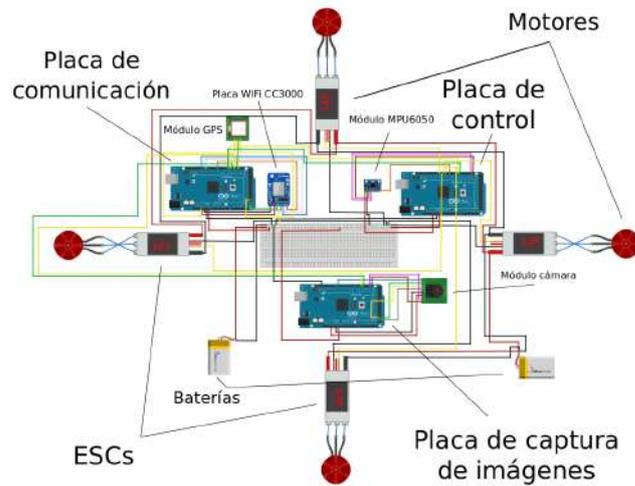


Fig. 3. Esquemático de las conexiones del dron.

función de realizar las capturas, recoger la imagen comprimida y enviarla hasta la placa Arduino de comunicación. Esto es debido a que la cámara requiere de mucha alimentación, así como la placa WiFi, de modo que la mejor opción era utilizar placas independientes. Pues la operación de tomar capturas también requiere un gran uso de la memoria así como tiempo de ejecución y procesamiento bloqueante, por lo que interfería con las operaciones principales del resto de placas.

#### IV. ESQUEMA GENERAL, CONTROL Y COMUNICACIÓN

**P**ARA la comunicación entre los operadores de control y los drones se ha establecido una arquitectura completa de esquema cliente-servidor, donde un servidor de gestión realizará la función de intermediario entre los clientes móviles y los drones, consiguiendo así una funcionalidad mucho más flexible, escalable y configurable que en una arquitectura punto a punto entre clientes y drones.

Del mismo modo al introducir un servidor de gestión ejecutándose en una máquina más potente como podría ser un computador se permiten realizar funciones más sofisticadas, como el reconocimiento de objetos.

El esquema general de la arquitectura de comunicación entre clientes, servidor y drones queda reflejado en la Figura 4. Como se puede apreciar, aparecen 5 componentes en esta arquitectura:

##### 1) Drones

Este elemento representa a los posibles drones. Estos se conectan al servidor de gestión a través de un puerto propio y a partir de ese momento esperan las operaciones que le lleguen por parte del cliente a través del servidor de gestión o bien desde el propio servidor.

##### 2) Clientes móviles

Los clientes móviles ejecutando la aplicación en Android se conectarán al servidor a través de un puerto distinto al de los drones, y a partir de este momento podrán visualizar en la aplicación los drones disponibles que puedan controlar, de modo que una vez se establezca la conexión indirecta entre dron o grupo de drones y cliente (de forma indirecta debido a que la información siempre viaja a través del servidor de gestión) ya se podrán controlar.

##### 3) Servidor de gestión

Es el elemento principal de esta arquitectura, se encarga de hacer de enlace entre el resto de componentes. El servidor de gestión se convierte tanto en servidor para los clientes móviles como para los drones, de modo que este guarda todas las conexiones existentes y permite establecer la comunicación entre todos los clientes. Del mismo modo actúa de pasarela para los servidores de reconocimiento de objetos y para el simulador de DroneKit [6].

##### 4) Servidor de reconocimiento de objetos

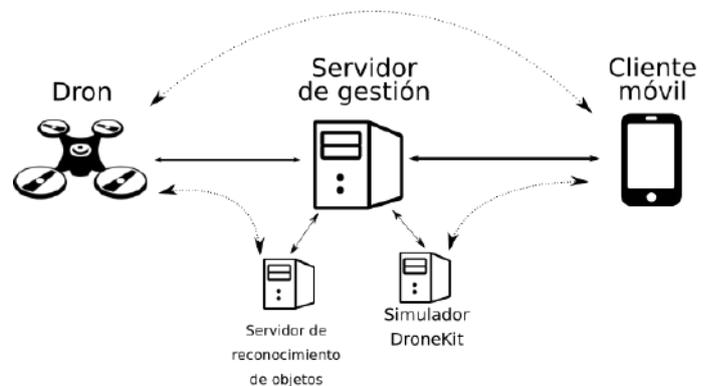


Fig. 4. Esquema final para la arquitectura cliente-servidor

Este servidor (ejecutándose en la misma máquina que el servidor de gestión o no), se encargará de recoger las imágenes procedente de los drones, hacer un análisis de los objetos presentes en dicha imagen y devolver la información al servidor de gestión, que se encargará de realizar las operaciones oportunas con dicha información.

## 5) Simulador DroneKit

Para verificar el funcionamiento de las aplicaciones que se desarrollen para el sistema, se cuenta con un servidor de simulación de vuelo de drones. Es un servidor que contiene al simulador de DroneKit (simulador de drones que utilizan el protocolo *MAVLink*), que permite a los clientes móviles conectarse a los drones simulados para hacer pruebas de forma segura. Así se puede comprobar la ejecución de las órdenes de control en los dispositivos simulados antes de realizarlas en un dron real, reduciendo el riesgo de accidente.

## V. PROTOCOLO DE COMUNICACIÓN

**P**ARA comunicar las distintas partes ha sido necesario el diseño de un protocolo de comunicación propio que permitiese realizar todas las operaciones deseadas a la vez que tenía en cuenta las limitaciones existentes en los componentes. Partiendo de las bases establecidas, se obtiene un protocolo sencillo y con paquetes y mensajes muy sencillos que realizan la comunicación entre las distintas partes de forma ligera y fluida. El formato de los mensajes definidos está representado en la Figura 5.

## VI. APLICACIÓN ANDROID PARA EL CONTROL DE DRONES

**P**ARA el control del dron se ha elegido el desarrollo de una aplicación Android, ya que es una de las plataformas móviles con más dispositivos a nivel mundial, formando parte del 76.53% de los dispositivos móviles frente al 18.97% de iOS, el sistema operativo de Apple. Al tratarse de un dron controlado por WiFi, un dispositivo móvil se convierte en la herramienta más versátil, flexible y útil con la que controlar el dron debido a su masificación y comodidad de uso.

La aplicación desarrollada permite conectarse al servidor introduciendo un usuario y contraseña válida o bien registrando la cuenta por primera vez. Una vez completado el login debe aparecer una lista con los drones disponibles, de modo que el cliente pueda seleccionar alguno de ellos y comenzar el control. El control se lleva a cabo utilizando el acelerómetro y el giroscopio del dispositivo móvil. Permite además visualizar la localización del dron en un mapa gracias a los datos de longitud y latitud obtenidos del dron



Fig. 5. Formato de los mensajes

y del mismo modo es capaz de planear rutas y seleccionar una localización destino para los drones simulados mediante DroneKit. También es posible mediante la aplicación visualizar las imágenes o capturas realizadas por el dron, las cuales son enviadas en un primer lugar al servidor de gestión por este y finalmente hasta la aplicación móvil.

De este modo la aplicación se basa en 4 ventanas (*activities* o *actividades*, según la nomenclatura de Android). El *activity* inicial es una ventana de login, donde el usuario puede registrarse por primera vez o conectarse con una cuenta existente válida, en un segundo *activity* se muestran los drones disponibles, un tercer *activity* sirve para la calibración de los sensores y asegurar que el dispositivo se encuentra en la posición correcta y por último en el cuarto *activity* se lleva a cabo el control del dron. En la Figura 6 se muestra una captura de esta última ventana donde se realiza el control del dron.

## VII. SERVIDOR DE GESTIÓN

**E**N un primer momento la comunicación entre cliente móvil y dron iba a ser directa, sin embargo la implantación de un servidor de gestión y por lo tanto una arquitectura cliente-servidor tiene un gran número de ventajas sobre la implementación inicial.

En la figura 7 se muestra el esquema que se ha utilizado para el diseño y la implementación del servidor de gestión. Este servidor de gestión o servidor principal está formado por 4 bloques funcionales, reflejados en la figura mediante diferentes colores.

- 1) Base de datos, marcada en rojo.
- 2) Servidor central, marcado en naranja.
- 3) Implementación del simulador Dronekit, marcado en verde.
- 4) Servidor de reconocimiento de objetos, marcado en azul.

## VIII. PRESUPUESTO

**E**L presupuesto derivado de la compra de los diferentes productos, elementos y dispositivos que conforman el dron queda reflejado en la tabla I.

Como se puede observar, los elementos más costosos son los que forman parte de la propia estructura y elementos básicos del dron, como motores, ESC o baterías. Por otro

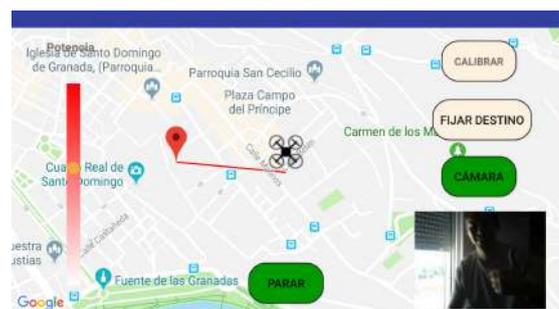


Fig. 6. Activity de control del dron en la aplicación Android

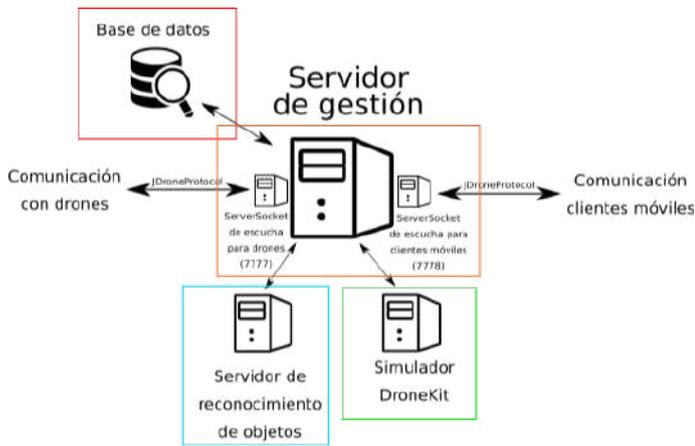


Fig. 7. Arquitectura para el diseño del servidor de gestión

lado, los elementos electrónicos ocupan un segundo lugar, pues como se ha visto uno de los objetivos del proyecto era conseguir un dron completamente funcional con todas las características posibles y con un precio lo más reducido posible.

### IX. CONCLUSIONES

EN este proyecto se ha descrito y llevado a cabo la construcción de un dron desde cero utilizando como hardware de control Arduinos y los componentes esenciales compatibles con esta tecnología. Para ello, se han descrito además todas las partes que componen un cuadricóptero, y cómo funciona este.

Para que fuera un sistema versátil y programable, el control se realiza a través de WiFi utilizando para ello un dispositivo móvil. El dispositivo móvil se comunica con el dron a través de un servidor de gestión.

El dron incorpora una cámara, por lo que además es posible transmitir las capturas tomadas hasta el servidor, que dispone de un servidor de detección de objetos. Este servidor consigue analizar la información contenida en dicha imagen y extraer los datos de ella.

Por otro lado, el servidor cuenta con otro módulo que permite comunicarlo con un simulador de *Dronekit*, de modo que se demuestra cómo gracias al diseño del servidor de gestión, el sistema es capaz de integrar otras tecnologías sin

tener que cambiar el código del dron o del cliente.

Personalmente ha sido un proyecto arduo y difícil, que ha requerido una enorme cantidad de horas, y en el que ha sido necesario un gran trabajo de investigación y experimentación. Uno de los aspectos más complicados ha sido la calibración del dron, ya que era necesario conocer con detalle el funcionamiento de este. Del mismo modo, al tratarse de componentes relativamente baratos, se imponen las limitaciones de estos, por lo que es necesaria la resolución de una gran cantidad de problemas para integrar todos los dispositivos, y que el funcionamiento fuera correcto.

Sin embargo, a pesar de las limitaciones, ha sido un proyecto muy completo, motivador y con gran carga experimental. Se integra tanto electrónica como telemática en él, de modo que ha servido para afianzar y poner a prueba los conocimientos adquiridos a lo largo de la carrera. Del mismo modo he necesitado ser capaz de resolver errores y problemas muy concretos y para los cuales apenas existe documentación, y ante ellos ofrecer y establecer soluciones válidas y razonables.

### AGRADECIMIENTOS

En primer lugar quisiera agradecer este proyecto y junto a él todo lo vivido durante este grado a mis padres, pues sin su trabajo y dedicación no habría sido posible que yo pueda estar escribiendo estas líneas. Siempre me faltarán palabras de agradecimiento para ellos, pues creo que no hay manera correcta de expresarlo como se merecen.

También me gustaría agradecer estos años a mis hermanos, Antonio Lara Sarabia y Ángel Lara Sarabia, pues sin duda alguna son las personas que nunca me han fallado y que siempre han estado ahí para prestarme un hombro donde apoyarme.

Por último agradecer a todos los profesores y docentes que durante estos años se han preocupado por enseñarnos nuevos conocimientos y hacernos crecer como personas. Y entre ellos me gustaría dedicar un agradecimiento especial a mi profesor del instituto Pedro Jesús Ramón Torregrosa, que tanto me ha ayudado en este proyecto y que desde un primer momento impulsó y apoyó como nadie, uno de esos profesores que nunca te pone una barrera y que no solamente te enseña conocimientos, sino a disfrutar aprendiendo.

### REFERENCIAS

- [1] Ardupilot. Disponible: <http://ardupilot.org/>
- [2] MAVLink: protocolo de comunicación para drones. Autor: Ernesto Santana. Disponible: <http://www.xdrones.es/mavlink/>
- [3] Arduino. Disponible: <https://www.arduino.cc/>
- [4] CC3000 IEEE 802.11b/g Texas Instruments. Disponible: <http://www.ti.com/lit/ds/symlink/cc3000.pdf>
- [5] Módulo acelerómetro y giroscopio MPU6050. Disponible: <https://playground.arduino.cc/Main/MPU-6050/>
- [6] Dronekit. Disponible: <http://python.dronekit.io/>

Tabla I  
PRESUPUESTO TOTAL DEL DRON

Componente	Precio por unidad	Número de unidades	Precio total
Motor FC 28-22	6.86€	4	27.44€
ESC Hobbyking 12A blueseries	7.88€	4	31.52€
Batería 3700mAh 3s 25C	21.28€	2	42.56€
Hélices	4.3€	1	4.3€
Chasis	5€	1	5€
Cámara ArducamOV2640	22.27€	1	22.27€
WiFi Shield CC3000	22.89€	1	22.89€
Módulo MPU6050	9.85€	1	9.85€
Barómetro BMP180	3.95€	1	3.95€
Arduino Mega 2560 marca Elegoo	7.5€	3	22.5€
Cableado	5.95€	1	5.95€
Cargador para baterías 3S Hobbyking B3AC	6.75€	1	6.75€
Comprobador de baterías	2.56€	2	5.12€
		<b>Total:</b>	<b>210.1€</b>

# Herramienta para la importación de CVN a la plataforma SICA

Autor: Juan Carlos Angulo Santos, e-mail: jcangulosantos@gmail.com  
Tutor: Jorge Navarro Ortiz, e-mail: jorgenavarro@ugr.es  
Titulación: Máster en Ingeniería de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—En la actualidad, los investigadores deben incluir su Currículum Vitae en diferentes plataformas y formatos para distintas convocatorias y entidades. Por ello, el Ministerio de Economía, Industria y Competitividad propone el uso del Currículum Vitae Normalizado (CVN) para muchas de sus convocatorias, ya que define un mismo formato de presentación de los datos curriculares.

En el caso de Andalucía, los investigadores se incluyen en grupos de investigación del Sistema de Información Científico Andaluz (SICA). Sin embargo, aunque la plataforma SICA permite la exportación al formato CVN, no implementa la posibilidad de importación del mismo. Este Trabajo de Fin de Máster pretende abordar esta carencia, permitiendo así que un investigador que tenga generado un CVN pueda importar sus datos a SICA de forma automatizada, en lugar de introducir nuevamente todos los elementos que incluye dicho documento.

**Palabras clave**—Análisis de tráfico, CV, CVN, Currículum Vitae Normalizado, Desarrollo Web, Investigadores, Importación

## I. INTRODUCCIÓN

EN la actualidad, los investigadores deben incluir su Currículum Vitae (CV) en diferentes plataformas y formatos para distintas convocatorias y entidades. Así, por ejemplo, el Ministerio de Economía, Industria y Competitividad propone el uso de un formato de CV conocido como Currículum Vitae Normalizado (CVN) [1], perteneciente a la Fundación Española para la Ciencia y la Tecnología (FECYT) [2], para muchas de sus convocatorias. Esta entidad proporciona un editor para que cada investigador gestione y genere dicho CVN a través de una dirección Web.

Como se detallará más adelante, el CVN, es una norma estándar que define un mismo formato de presentación de los datos curriculares de los investigadores y permite que estos puedan incluirse en las diferentes bases de datos de las instituciones. En definitiva, sirve para que los investigadores puedan presentar su CV en un formato digital único y normalizado, lo que les permite actualizar de forma automática todos sus datos, sin tener que volver a introducirlos de nuevo.

En el caso de Andalucía, los investigadores se incluyen en Grupos de Investigación del Sistema de Información Científico de Andalucía (SICA) [3], una aplicación de la Secretaría General de Universidades, Investigación y Tecnología adscrito a la Agencia Andaluza de Evaluación de la Calidad y la Acreditación Universitaria, cuyo objetivo principal es recopilar y ofrecer información sobre la actividad científica e investigadora de las Universidades Andaluzas [4].

Sin embargo, en SICA no existe ninguna opción para la importación de ficheros, incluido el CVN, lo que resulta en un esfuerzo adicional para los investigadores que ya posean sus datos en dicho formato, el volver a introducir los mismos elementos en SICA. La única opción para introducir elementos en SICA es a través de la interacción del usuario con la plataforma, añadiéndolos manualmente cada elemento.

Este Trabajo de Fin de Máster pretende abordar esta carencia, permitiendo así que un investigador que tenga generado un CVN, con cualquier editor existente, pueda importar dichos datos a SICA.

Por lo tanto, el objetivo del proyecto será desarrollar una herramienta con la capacidad de importar elementos incluidos en el CVN a la plataforma SICA de forma automática. Para ello se permitirá que un usuario introduzca un documento CVN en la herramienta, extrayendo todos los posibles datos curriculares que se incluyen en el mismo. Tras esto, será posible seleccionar los elementos que desean ser importados a SICA, mostrando la mayor cantidad de información posible obtenida del CVN.

## II. ESTADO DEL ARTE

En esta sección se desarrolla el estado del arte, dando una perspectiva global del área con la que se trabaja en el proyecto.

En el caso de la plataforma SICA a la que se atañe este proyecto, la importación de un documento en formato CVN no es posible debido a su no implementación ni certificación por parte de FECYT. Sin embargo, existen diferentes plataformas que sí permiten la importación de un fichero CVN para registrar todos los datos. Actualmente, existen 122 instituciones certificadas, de las cuales todas permiten la exportación al formato CVN, pero solo 15 de ellas permiten la importación.

Por otro lado, el editor CVN se trata de una aplicación online que pone a disposición de los investigadores la posibilidad de crear un currículum vitae desde cero o modificar el existente. En este editor es posible tanto la exportación como la importación al formato CVN.

Numerosas plataformas dedicadas a la investigación e introducir reseñas curriculares como Web Of Science (WOS) o Scopus ya trabajan con el formato CVN, permitiendo exportar los datos a dicho formato. Sin embargo, aún no permiten la importación del formato CVN.

Existen otro tipo de herramientas que tienen como finalidad importar datos curriculares, por lo general, de forma manual. No existen demasiadas que permitan la importación de un

CV, ya que este proceso requiere que el CV tenga un formato específico, como es el caso del CVN, y que sus campos estén adaptados a la plataforma a la que se requiere importar. Es el caso de Europass [5], que consta de un CV con un formato propio, que además puede ser importado en la plataforma de Europass y que es promocionado por SEPIE (Servicio Español Para la Internacionalización de la Educación).

Otro conocido estándar utilizado para el intercambio de información científica es el esquema METS [6], utilizado para la codificación e electrónica de un documento u objeto dentro de una biblioteca digital.

Por otro lado, existen otro tipo de formatos de plataformas y formatos de CV ajenos a España y la UE, como es el caso de CVar del Ministerio de Ciencia, Tecnología e Innovación Productiva de Argentina [7].

Como se ha podido observar en el estado del arte de este proyecto, la mayoría de plataformas dedicadas a la investigación y reseñas curriculares no permiten importación de un Currículum Vitae, sino que están diseñadas para poder importar citas o referencias bibliográficas. Es por eso, que es necesario un formato de normalización de CV como es el caso de CVN para implementar en diferentes plataformas la funcionalidad de importar los datos almacenados en un CV.

### III. REQUISITOS

En esta sección se llevará a cabo un análisis del proceso de diseño realizado para este proyecto, incluyendo, de la manera más completa posible, los requisitos relacionados con la funcionalidad y las decisiones previas que se deben cumplir.

A continuación se listarán los requisitos mínimos que serán necesarios para el funcionamiento del proyecto, y que serán útiles a la hora de tener presente las ideas del diseño:

- **Obtención del CVN-XML incrustado en el CVN-PDF:** Será necesario que la herramienta realizada sea capaz de obtener el XML incrustado en el CVN-PDF, en dónde se encuentra toda la información curricular.
- **Análisis de la información del CVN-XML:** La herramienta deberá poder analizar los distintos campos de los que se compone el currículum en el formato CVN y mostrar por pantalla los que sean compatibles con los elementos de CV incluidos en la plataforma SICA.
- **Análisis de campos y tráfico HTTPS de SICA:** Será necesario conocer todos los posibles campos de los elementos que pueden ser añadidos a SICA, observando qué valores de los mismos pueden ser obtenidos del CVN. Adicionalmente, será necesario analizar el tráfico a la hora de añadir de un elemento en la plataforma SICA para realizar la importación de un CVN. Este tráfico se utilizará para añadir de manera automática a SICA los datos curriculares del CVN.
- **Importación de CVN-PDF o CVN-XML a SICA:** Será necesario desarrollar una herramienta con la capacidad de importar un documento CVN a la plataforma SICA.
- **Utilidad y facilidad:** Uno de los factores más importantes en la herramienta a desarrollar debe ser su utilidad para los distintos investigadores que la utilicen..
- **Interfaz sencilla e intuitiva:** La interfaz de la herramienta debe ser sencilla e intuitiva de modo que no sea complicado utilizarla por los usuarios que lo deseen.

- **Herramienta tolerante a fallos:** La herramienta desarrollada debe ser tolerantes a fallos e informar al usuario de posibles errores.

Además de estos requisitos, una decisión previa al diseño ha sido desarrollar la herramienta a través de una interfaz web, por lo que será necesario utilizar lenguaje HTML, CSS y JavaScript para el lado del cliente y para la parte del servidor en donde se utilizará lenguaje PHP. Esta decisión se debe a que una herramienta desarrolla sobre una interfaz web ofrece un mayor acceso a cualquier usuario que desee importar su CVN.

### IV. PLANIFICACIÓN Y PRESUPUESTO

En esta sección se describirá la planificación que se ha adoptado para la realización del proyecto, y se presentará una aproximación del presupuesto total necesario para su llevada a cabo.

#### A. Planificación

A continuación, se enumeran las tareas realizadas, en orden cronológico:

- Tarea 1: Revisión bibliográfica
- Tarea 2: Definición de requisitos
- Tarea 3: Aprendizaje sobre el formato CVN
- Tarea 4: Aprendizaje sobre la plataforma SICA
- Tarea 5: Diseño del proyecto
- Tarea 6: Análisis de todos los nodos del CVN
- Tarea 7: Análisis de los elementos y campos en SICA
- Tarea 8: Análisis del tráfico en SICA
- Tarea 9: Herramienta de importación de CVN a SICA
- Tarea 10: Fase de pruebas
- Tarea 11: Memoria técnica del proyecto
- Tarea 12: Exposición del Trabajo de Fin de Máster

#### B. Presupuesto

En este apartado se tratará el tema del presupuesto del proyecto, en el que se listarán los recursos utilizados y el precio estimado correspondiente a cada uno.

1) *Recursos humanos:* Los recursos humanos son aquellos en los que ha sido necesaria la intervención humana para la realización del proyecto. En el caso del presente trabajo los recursos humanos son:

- Juan Carlos Angulo Santos: Autor del proyecto, graduado en Ingeniería de Telecomunicaciones y alumno del Máster en Ingeniería de Telecomunicaciones.
- Jorge Navarro Ortiz: Tutor del proyecto y profesor contratado doctor de la Universidad de Granada del Departamento de Teoría de la Señal, Telemática y Comunicaciones.

Se ha estimado que el coste de un ingeniero graduado en ingeniería de tecnologías de telecomunicación es de aproximadamente 25 euros/hora y el de un doctor de la Universidad de Granada es de aproximadamente 50 euros/hora. Con estos datos es posible hacer un cálculo estimado de los recursos humanos invertidos en el proyecto. Aproximadamente, el tutor del proyecto, entre tutorías y revisión del trabajo, invertirá 15 horas.

En la siguiente tabla (véase tabla I) podemos observar el número de horas dedicado y precio estimado para cada tarea.

Tarea	Tiempo (Horas)	Coste (Euros)
Revisión bibliográfica	20	500
Definición de requisitos y elección de herramientas/plataformas	20	500
Aprendizaje sobre el formato CVN	30	750
Aprendizaje sobre la plataforma SICA	30	750
Diseño del proyecto	40	1000
Análisis de todos los posibles campos del formato CVN	50	1250
Análisis de los elementos y campos en SICA	70	1750
Análisis del tráfico en SICA	70	1750
Creación de herramienta de importación de CVN a SICA	90	2250
Fase de pruebas	30	750
Memoria técnica del proyecto	50	1250
Exposición del TFM	25	625
Trabajo de ingeniero doctor	15	375
<b>Total</b>	<b>540</b>	<b>13500</b>

Tabla I  
ESTUDIO DE COSTES PARA LOS RECURSOS HUMANOS.

Como se observa, según la estimación realizada, el presupuesto total para los recursos humanos será de 13500 euros.

2) *Recursos hardware*: Respecto a los recursos hardware utilizados para este proyecto, únicamente ha sido necesario un Ordenador portátil ASUS R510, con procesador Intel Corte i7 6700 HD y 12 GB de memoria RAM, con un coste de 600 euros. Es importante un buen equipamiento informático que permita la ejecución del proyecto en un tiempo apropiado, aumentando la productividad a la hora de trabajar.

Considerando que la vida útil de un portátil es de cinco años, y dado que este proyecto ha tenido una duración aproximada de un año, el presupuesto equivalente del portátil utilizado será de aproximadamente 120 euros.

3) *Recursos software*: Estos son los recursos software que se han utilizado:

- Sublime Text: Es un editor de texto gratuito dedicado a código, utilizado en este proyecto para la programación en PHP. Además, se ha utilizado para visualizar los archivos CVN-XML obtenidos del CVN-PDF.
- XAMPP: Es un servidor independiente de plataforma de software libre que se ha utilizado para emular el servidor en una página web y hacer distintas pruebas.
- TeXstudio: Es un editor de LaTeX de código abierto que se ha utilizado para redactar la memoria del proyecto.
- GanttProject: Es el programa utilizado para realizar el diagrama de Gantt que se mostraba anteriormente.
- SourceTree: Es un cliente GUI utilizado para manejar repositorios GIT que se ha empleado para realizar copias de seguridad del proyecto. Además, permite de forma sencilla observar los cambios producidos en cada versión que se ha subido.
- Mitmproxy: Se trata de un proxy *Man in the middle*, utilizado para examinar el tráfico HTTPs al generar elementos en SICA. Se ha utilizado una versión para web conocida como mitmweb.

En el caso de los recursos software, se ha buscado que todos los programas utilizados tuviesen un coste gratuito, con el propósito de hacer este proyecto lo más económico posible.

4) *Presupuesto final*: El presupuesto final será la suma del estudio de costes realizado para los recursos humanos, hardware y software. Puede verse el precio estimado para cada sección en la tabla II:

Recursos	Presupuesto (Euros)
Recursos humanos	13500
Recursos hardware	120
Recursos software	0
<b>Total</b>	<b>13620</b>

Tabla II  
PRESUPUESTO FINAL DEL PROYECTO.

En definitiva, el presupuesto total del proyecto será de 13620 euros, considerando todos los aspectos anteriores.

## V. CURRÍCULUM VITAE NORMALIZADO

En esta sección se detallará qué es el estándar CVN y se describirá la norma CVN que incluye las especificaciones técnicas del formato.

### A. El estándar CVN

El Currículum Vitae normalizado [1], conocido por sus siglas CVN, es un proyecto de la Federación Española para la Ciencia y la Tecnología (FECYT) y del Ministerio de Economía, Industria y Competitividad del gobierno español, para la implantación de las tecnologías de la información en la gestión de la actividad investigadora, con el objetivo de evitar que el investigador tenga que completar de nuevo su CV en los formularios o bases de datos de las convocatorias. Este proyecto comenzó a implantarse en 2006 en universidades y organismos de investigación, adaptando la información ya cargada de los investigadores al estándar CVN.

El CVN es una Norma estándar que define un mismo formato de presentación de los datos curriculares de los investigadores y que posibilita la interoperabilidad con las diferentes bases de datos de las instituciones. Tiene como objetivo que los investigadores puedan presentar su currículum en un formato digital único y común que les permite actualizar de forma automática su información curricular, sin tener que volver a introducirla de nuevo. Por lo tanto, CVN facilita la presentación del CV de una forma unificada en las convocatorias de ayudas públicas del Plan Estatal y autonómicas.

Desde el punto de vista tecnológico, un currículum vitae en formato CVN es un fichero electrónico en PDF (CVN-PDF) con tecnología CVN-XML incrustada, de tal manera que la información normalizada que contiene puede ser tratada electrónicamente por distintas bases de datos; con la intención de evitar que los investigadores tengan que teclear su currículum en los distintos formularios o aplicaciones que se utilizan para la gestión curricular y en las diferentes convocatorias de ayudas.

Este fichero en formato CVN-PDF puede ser archivado, visualizado o impreso por el investigador, pero a efectos de seguridad e integridad de la información, cualquier modificación o manipulación posterior a su obtención desde una aplicación informática lo invalidará. Por lo tanto, cualquier modificación o subsanación de error que se desee, deberá realizarse previamente en la aplicación curricular de una institución o a través del editor CVN de FECYT.

## B. La norma CVN

La norma CVN es el conjunto de especificaciones técnicas que regulan todos los campos del currículum estándar, sin tener ninguno de ellos un carácter obligatorio. Por lo tanto, es el reglamento que deben seguir todos los currículos adaptados al formato CVN para que tengan cohesión en un único formato. Esta Norma comprende la información que un investigador, tecnólogo o innovador puede necesitar para reseñar su trayectoria e incluye su interacción con otras dedicaciones.

La Norma CVN pretende sentar las bases para la creación de una red de intereses curriculares comunes para el Sistema Español de Ciencia Tecnología e Innovación. Con este objetivo, la FECYT ha elaborado un documento normativo para la estandarización de los Currículos Vítae de I+D+i.

En la figura 1 se observa un ejemplo de un CVN-XML en donde se muestra el formato que sigue:

Fig. 1. Ejemplo de formato CVN-XML.

CVN-XML es un modelo abstracto con la capacidad de describir cualquier tipo de actividad profesional, actuando como un traductor universal y un contenedor de diferentes CVs. Un conjunto de lenguajes XSD [8] y valores XML proporcionan toda la semántica y síntesis necesaria para aplicar el modelo CVN-XML.

Esta Norma es de carácter privada y FECYT únicamente se explica y detalla todo el proceso tanto desde el punto de vista funcional como tecnológico del proyecto a aquellas instituciones que estén certificadas. Esto quiere decir que la información sobre el formato de la tecnología CVN-XML es privada y pertenece a FECYT, por lo que no existe información pública sobre el formato y ha sido necesario investigar todo este proceso por cuenta propia sin obtener ninguna información adicional, lo que ha dificultado el desarrollo del presente trabajo.

Como se observa en la figura 1 el formato CVN se compone de diversos campos, que se estructuran de una forma ordenada. Cada *item* o elemento curricular que puede encontrarse en el CVN tiene unos atributos utilizados para la descripción de cada campo.

## VI. DISEÑO E IMPLEMENTACIÓN

En esta sección se abordan los aspectos referentes al diseño y a la implementación de la herramienta que se pretende desarrollar en este proyecto, cuyos términos teóricos se han desarrollado anteriormente.

## A. Diseño

1) *Lenguajes utilizados*: Para el desarrollo de la herramienta encargada de la importación de CVN a la plataforma SICA se ha utilizado el lenguaje PHP [9], en su última versión (7.19) que es estable y corrige errores de seguridad de versiones anteriores. PHP es un lenguaje de código abierto adecuado para el desarrollo web y que puede ser incrustado en HTML.

Para la parte del cliente, es decir, lo que se visualizará por pantalla al usuario, se han utilizado los conocidos lenguajes HTML5, JavaScript y CSS3 [10]. El lenguaje HTML aporta la estructura básica de los sitios, mejorada por otras tecnologías como CSS y JavaScript.

2) *Man in the middle proxy (mitmproxy)*: Para el análisis del tráfico al generar elementos en SICA se ha utilizado una herramienta conocida como mitmproxy [11], un proxy *man in the middle* interactivo para HTTP y HTTPS con una interfaz por consola. También se ha utilizado otra de sus aplicaciones, conocida como mitmweb, que posee una interfaz basada en web para mitmproxy. Este análisis de tráfico será utilizado para posteriormente poder replicar el tráfico necesario para añadir elementos a SICA de forma automática.

Las siglas de MITM (Man-In-The-Middle) hacen referencia al proceso utilizado para interceptar e inferir paquetes de datos teóricamente opacos. La idea básica es pretender ser el servidor para el cliente y pretender ser el cliente para el servidor, mientras que la herramienta se sitúa en el medio de ambos decodificando el tráfico proveniente de ambos lados.

## B. Implementación

Para la implementación de este proyecto se han seguido tres pasos: análisis del formato CVN, análisis del tráfico en SICA y el desarrollo de la herramienta.

1) *Análisis del formato CVN*: Para realizar en análisis del formato CVN se obtuvo inicialmente un fichero CVN-PDF con diversos elementos de un CV, como pudieran ser la situación profesional, participación en congresos o artículos publicados. En esta primera toma de contacto se investigó como se podía obtener el XML incrustado en el PDF, ya que como se comentaba anteriormente, esta información no es pública. Tras este proceso de investigación se consiguió finalmente obtener un archivo CVN-XML que se encontraba incrustado en un CVN-PDF. Este CVN-XML se encontraba comprimido en un archivo que se obtenía de una cadena de caracteres codificada en Base64 e incrustada en el código fuente del fichero CVN-PDF.

Por lo tanto, una vez obtenido el fichero CVN-XML es posible analizar todos los componentes y estructura del archivo. En primer lugar, era necesario identificar los nodos del CVN-XML más importantes y conocer a que hacen referencia. En un CVN existen tres nodos principales (*Version*, *Agent* y *CvnItem*) de los que cuelga información de diferentes tipos en forma de nodo. De esos tres nodos, el *CvnItem* es el que contiene la información curricular, y existe un nodo de ese tipo por cada elemento que contenga el CV. Por lo tanto, ese es el nodo que interesa para la realización de este proyecto, ya que ahí se encuentran los datos del CV que hacen referencia, por ejemplo, al título del elemento o a la categoría a la que pertenece.

Además, cada nodo contiene un identificador del tipo de elemento incluido en el CV. Este identificador hace referencia a un listado de secciones que propone FECYT y que identifican al elemento, como por ejemplo, la situación laboral actual o publicaciones en congresos.

Sabiendo esto, es posible analizar todos los posibles campos que puede incluir cada tipo de elemento en un CVN según su identificador, como por ejemplo lugares, fechas o entidades. Cada posible elemento, distinguido por su identificador, posee unos campos determinados, por lo que es importante intentar encontrar todos los campos que puede contener un determinado nodo y compararlos con los existentes en la plataforma SICA.

Por lo tanto, es necesario obtener todos los posibles elementos existentes en SICA, y obtener los campos encontrados para cada uno de ellos. Cabe destacar que existen ocasiones en el que un elemento de SICA contiene información en su página web que no se ve reflejada en el CVN-XML, y a que no existen nodos para contener esa información, por lo que no todos los campos de un CVN son importables a SICA.

2) *Análisis del tráfico en SICA*: Para conseguir añadir elementos en SICA a partir de la herramienta que se programará para este proyecto será necesario analizar el tráfico que se genera al introducir un elemento en la página web de la plataforma. Este análisis se utilizará para enviar peticiones a SICA de manera automática, con el objetivo de añadir los elementos que contiene un CVN. Previamente al análisis del tráfico es necesario analizar cómo está estructurada la plataforma SICA y qué elementos pueden ser añadidos en ella.

SICA hace uso de diferentes tipos de contenido, que son: Publicación en revista, Capítulo de libro, Libro, Aportaciones a congreso y Otros. Cada uno de estos tipos de contenido tienen su forma particular de ser añadidos en SICA, por lo que no se sigue un patrón estándar a la hora de generar el elemento. Los cuatro primeros tipos que se han nombrado permiten añadir un elemento curricular de ese mismo tipo, siguiendo cada uno una estructura determinada, siempre con los mismos campos. Sin embargo, la categoría Otros varía respecto a las anteriores, ya que permite seleccionar entre un numeroso conjunto de tipos de elementos y subtipos que añadir, y cada uno de ellos sigue una estructura diferente, cambiando continuamente los campos a mostrar.

Una vez conocida la estructura de la página web de SICA y cómo se añaden elementos, se procedió al análisis del tráfico, que se ha realizado con la herramienta mitmproxy. Esta herramienta, explicada en VI-A2, que permitirá analizar todas las peticiones enviadas desde un navegador y las respuestas recibidas desde el servidor. En las peticiones que se envíen se podrán observar el nombre y el valor de todas las variables que son enviadas a través de un formulario y a qué página web, y en las respuestas se podrá observar el código HTML que se recibe tras la petición.

Con esta herramienta se ha conseguido obtener la información necesaria que hay que enviar a través de peticiones para conseguir añadir todos los posibles elementos existentes en SICA. Con esta información del tráfico y la obtenida del CVN-XML es posible importar todos aquellos elementos encontrados en el CV a la plataforma SICA.

3) *Desarrollo de la herramienta*: El desarrollo de la herramienta se divide entre la realización de la parte cliente, que será la que visualice el usuario final, y la parte servidora, encargada de todo el procesamiento para conseguir importar un CVN a SICA.

En primer lugar, será necesaria una pantalla para introducir las credenciales de SICA, ya que no es posible añadir elementos en la plataforma sin iniciar sesión. Esta pantalla de inicio de sesión constará de dos campos para introducir el nombre de usuario y contraseña, además de un botón para continuar. Cabe destacar, que la herramienta no podrá usarse a menos que el usuario haya iniciado sesión, lo que estará controlado en el servidor.

Una vez que se ha iniciado sesión en la herramienta se redirigirá al usuario a una página para introducir su CVN en la herramienta. Esta página es muy sencilla, ya que únicamente hace uso de un botón para seleccionar un fichero CVN-PDF o CVN-XML y otro para comenzar la importación.

Tras la importación de un CVN-PDF o CVN-XML válido, comenzará el análisis de todos los datos curriculares contenidos en el fichero. Sin embargo, a la hora de continuar puede darse el caso de que no sea posible clasificar uno de los elementos encontrados en el CVN en una determinada categoría de SICA, por lo que será necesario resolver los conflictos encontrados.

Esto es debido a que en el editor de FECYT pueden ser insertados tantos elementos como posibles códigos de identificación puede haber en el formato CVN-XML. Sin embargo, para la plataforma SICA esto no ocurre de esa manera, ya que un mismo código de identificación de *item* del CVN puede corresponder a varias categorías, tipos de elemento o subtipos de elemento de SICA al mismo tiempo. Esto supone que sea necesaria una página para que el usuario indique a qué categoría, tipo o subtipo de SICA pertenece un elemento encontrado en el CVN.

En la figura 2 se puede observar un caso en el que un elemento puede pertenecer a una o más categorías en SICA.

Fig. 2. Resolución de conflictos para un elemento.

Por lo tanto, suponiendo que se encuentra un elemento en el CVN que pertenezca a dos o más categorías de SICA y no pueda ser clasificado automáticamente en alguna de ellas, se pasará a la página de pre-selección. En esta página se mostrarán los elementos del CVN que presentan conflictos y que deben ser clasificados en una categoría, tipo de elemento

o subtipo de elemento. Para ello, se mostrará el título del elemento encontrado en el CVN y el usuario deberá seleccionar a qué categoría de SICA pertenece.

Tras la resolución de conflictos se pasará a la página para seleccionar los elementos del CVN que se desean importar a SICA. Esta página muestra todos los elementos encontrados en el CVN, clasificados en categoría, tipo de elemento y subtipo de elemento de SICA, además de todos los posibles campos que puedan ser requeridos por la plataforma, tanto obligatorios como opcionales.

En la figura 3 se observa la página de selección de elementos sin desplegar ninguno de ellos, únicamente mostrando los títulos de los elementos curriculares encontrados. En la figura 4 se muestran todos los campos para un determinado elemento, con aquellos valores encontrados en el CVN.



Fig. 3. Página de selección de elementos obtenidos del CVN sin desplegar.

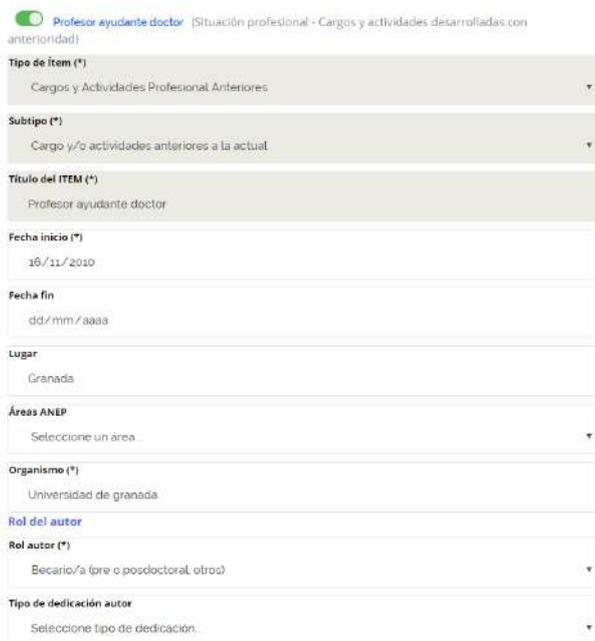


Fig. 4. Elemento desplegado en la página de selección.

Haciendo uso del análisis del fichero CVN-XML se extraerá la mayor cantidad de información posible para todos los elementos encontrados, mostrando al usuario aquellos campos que son compatibles con la plataforma SICA para cada elemento. Se le dará la opción al usuario de escoger aquellos

elementos que desea importar, marcando como obligatorios los campos que son requeridos por SICA.

Una vez el usuario escoge los elementos a importar deberá pulsar sobre el botón 'Importar', que llamará a la parte servidora encargada de realizar las peticiones necesarias para añadir dichos elementos a SICA.

## VII. CONCLUSIONES Y LÍNEAS FUTURAS

En esta sección se desarrollarán las conclusiones sacadas de este proyecto y las posibles líneas futuras que pueden ser requeridas o implementadas.

### A. Conclusiones

En el presente Trabajo de Fin de Máster se ha desarrollado una herramienta con la capacidad de analizar un fichero en formato CVN-PDF e importar a SICA la información obtenida del mismo.

El objetivo propuesto ha sido desarrollado con éxito, elaborándose una herramienta utilizada a través de una página web que actuará como analizador de ficheros CVN escogidos por el usuario y permitirá importarlos a SICA, mostrando todos los campos necesarios. Esto ha requerido realizar una parte dedicada al cliente, utilizando tecnología HTML, JavaScript y CSS, y una parte servidora en donde se realiza todo el procesamiento, como es el caso del análisis del CVN y la importación a SICA, utilizando para ello el lenguaje PHP.

### B. Líneas futuras

Aunque se dan por conseguidos los objetivos de este Trabajo de Fin de Máster, existen algunos aspectos que podrían mejorarse en un futuro, requerir un mantenimiento o implementarse adicionalmente.

Estos aspectos serán mantener actualizada toda la información de campos y del análisis de tráfico que han sido recopilados desde SICA, ya que la web podría cambiar su código HTML, añadir nuevos campos o alterar la forma en la que procesa los datos el servidor de la plataforma. Además podría desarrollarse la herramienta en otros lenguajes de programación o mejorar las funcionalidades de la misma, permitiendo una mayor interacción con SICA.

## REFERENCIAS

- [1] Ministerio de Economía, Industria y Competitividad. Currículo Normalizado (CVN). <https://cvn.fecyt.es/presentacion/que-es-cvn>, 2018.
- [2] Ministerio de Economía, Industria y Competitividad. Fundación Española para la Ciencia y la Tecnología. <https://www.fecyt.es/>, 2018.
- [3] Junta de Andalucía. Sistema de Información Científica de Andalucía (SICA). <https://sica2.cica.es>, 2018.
- [4] Junta de Andalucía. Centro Informático Científico de Andalucía (CICA). <https://www.cica.es/servicios/sistemas/sica2/>, 2018.
- [5] Europass. Europass. <https://europass.cedefop.europa.eu>, 2018.
- [6] Jerome P McDonough. METS: standardized encoding for digital library objects. *International journal on digital libraries*, 6(2):148–158, 2006.
- [7] Tecnología e Innovación Productiva de Argentina Ministerio de Ciencia. Sistema Integral de Gestión y Evaluación. <http://www.mincyt.gov.ar/accion/cvar-6467>, 2018.
- [8] Shudi Gao, C Michael Sperberg-McQueen, Henry S Thompson, Noah Mendelsohn, David Beech, and Murray Maloney. W3C XML schema definition language (XSD) 1.1 part 1: Structures. *W3C candidate recommendation*, 30(7.2):16, 2009.
- [9] PHP. ¿Qué es PHP? <http://php.net/manual/es/intro-what-is.php>, 2018.
- [10] Juan Diego Gauchat. *El gran libro de HTML5, CSS3 y Javascript*. Marcombo, 2012.
- [11] Maximilian Hils. Mitmproxy. <https://mitmproxy.readthedocs.io/en/v2.0.2/introduction.html>, 2018.

# Diseño e Implementación de algoritmos de planificación para 5G

Tutor: Juan Francisco Valenzuela Valdés; e-mail: juanvalenzuela@ugr.es

Tutor: Pablo Padilla de la Torre; e-mail: pablopadilla@ugr.es

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación

**Departamento de Teoría de la Señal, Telemática y Comunicaciones**  
**Universidad de Granada**

Autor: Alejandro Ramírez Arroyo, e-mail: alexrufox@correo.ugr.es

**Resumen**— El desarrollo de la tecnología es fundamental para poder satisfacer las demandas de tráfico de los usuarios. Entre los años 2017 y 2022, el crecimiento anual del tráfico mundial se estima en un 220%. Este crecimiento anual conduce a su vez a un aumento del número de usuarios conectados a redes IP, pasando de 2,4 a 3,6 dispositivos conectados por persona. Estos datos muestran la importancia del desarrollo de una infraestructura que pueda soportar esta carga de tráfico. Actualmente, las redes 4G son capaces de soportar esta carga. Sin embargo, es visible la necesidad de que surjan redes con tecnología 5G en un futuro próximo, siendo esperadas a partir de 2020. Por otro lado, es obligatorio reducir el consumo energético derivado del campo de las telecomunicaciones. En este proyecto se presenta un estudio de la optimización de varios parámetros del sistema 5G. Presenta la configuración de los diferentes parámetros básicos de un sistema, describe los diferentes planes de asignación de cada usuario a cada una de las estaciones base y optimiza la potencia transmitida por las estaciones base para obtener mejoras simultáneas en la capacidad del sistema, el consumo global de las estaciones base y el coste de la señalización en diferentes protocolos de movilidad. En otras palabras, por primera vez se presenta una optimización conjunta para varios niveles de red.

**Palabras clave**— Comunicaciones móviles, Redes heterogéneas, Optimización, Redes 5G, Eficiencia Energética

## I. INTRODUCCIÓN

Las necesidades de conexión a la red no paran de aumentar [1] [2]. El crecimiento de tráfico en las redes de comunicaciones es prácticamente exponencial y se está diseñando una nueva generación de comunicaciones móviles –5G– para responder a estas demandas. 5G aparece con la idea de tratar de cambiar la concepción de las redes que actualmente conocemos, multiplicando por cien el número de dispositivos conectados y ahorrando hasta un 90% el consumo de energía existente en la red entre sus principales novedades, lo que da lugar a mayor penetración y velocidad y menor latencia para el usuario, así como la introducción de IoT permitiendo las comunicaciones M2M (machine to machine) a mayor escala. Para ello, el término Massive MIMO [3] toma especial relevancia debido a que la puesta en escena de múltiples antenas tanto en transmisor como en

receptor permite multiplicar las velocidades obtenidas en tecnologías anteriores como 4G, obteniendo en diversos test capacidades de hasta 145 bits/s/Hz. Sin embargo, cuando todavía no se han lanzado tecnologías 5G ya se ha comenzado a hablar de que habrá algunas necesidades como el internet táctil que no serán satisfechas por las redes 5G.

Por otra parte, el consumo de energía en el ámbito de las telecomunicaciones no para de crecer y está empezando a tener impacto en la emisión global de gases de efecto invernadero [4]. Por lo tanto, la eficiencia energética (EE) es uno de los objetivos de diseño dentro de las futuras redes de comunicación. Una de las soluciones propuestas para hacer frente al aumento casi exponencial del tráfico móvil de datos es aumentar el número de estaciones base (EB). Por tanto, se hace necesario el ahorro de energía en las estaciones base, manteniendo las prestaciones en cuanto a capacidad de la red. Es por tanto que los operadores buscan aumentar simultáneamente la EE y la capacidad de la red, para ello surge una nueva métrica para evaluar el rendimiento de las estaciones base que son los "bits por julio". Es decir, cuantos bits podemos transmitir gastando un Julio. Este paradigma de diseño está sincronizado con las prioridades mundiales en materia de gestión de la energía, donde las cifras recientemente publicadas sugieren que el 3% [5] de la energía mundial es consumida por las infraestructuras TIC, induciendo alrededor del 2% [5] de las emisiones mundiales de CO<sub>2</sub>.

Pero no solo debemos tener en cuenta el tráfico de datos entre las EB y el usuario final ya que el tráfico que se genera desde la estación base hasta el router de acceso a internet cada día tiene mayor peso. Este tráfico tiene que ser encaminado de manera eficiente minimizando los costes de ruteado y señalización. Esto en el escenario planteado con una alta movilidad de los usuarios y la gran cantidad de estaciones base pequeñas –con un radio de cobertura pequeño– es una ardua tarea y por tanto será un coste más añadir a las redes de telecomunicación. Este coste resulta aún más importante cuando se ha visto que determinados costes, como los costes de señalización están creciendo a un ritmo un 50% más rápido que el tráfico de datos [6]. Por tanto, habrá que tener en cuenta tanto el consumo de las comunicaciones desde las EB a los router de acceso como la cantidad de datos de señalización usados por los distintos protocolos.

En consecuencia con todo lo anterior, los estándares 3GPP [7] actualmente están considerando nuevos enfoques de eficiencia energética en el diseño de las futuras redes móviles de 5G o más allá de las 5G. El enfoque se hace a través de toda la pila de protocolos examinando las ventajas y desventajas de forma global para todo el sistema buscando maximizar la nueva métrica “bits por julio” para todo el sistema y no solo para la capa física del mismo.

Para el desarrollo de un nuevo modelo de red es necesario una fase de planificación donde se deben tener en cuenta todos los parámetros necesarios para el correcto funcionamiento de la red, existiendo fases de prueba donde se lleva a cabo el testeo de la red. Además es necesaria una fase de optimización donde se estudia el funcionamiento del sistema observando como cambios en este pueden provocar cambios en parámetros vitales. La idea gira en torno a la minimización de costes (energéticos y de señalización) y la maximización de parámetros referentes al rendimiento como puede ser la capacidad por usuario o la capacidad global de la red.

En este proyecto se presenta un estudio de la optimización de varios parámetros del sistema 5G previa planificación de dicho sistema. La estructura de este documento es la siguiente. En la sección II se presenta la configuración del sistema, presentándose los diferentes modelos de cálculos de potencias, capacidades y costes de los distintos elementos que forman del sistema. A continuación, en la sección III, se describen los distintos planes de asignación de cada uno de los usuarios a cada una de las estaciones base sobre el terreno. En la sección IV, se muestra la optimización realizada a distintas variables del sistema tras aplicarle la asignación correspondiente. Por último, en la sección V, se extraen las conclusiones.

## II. CONFIGURACIÓN DE LOS PARÁMETROS DEL SISTEMA

Este apartado se va dividir en tres partes, a) configuración, b) cálculo de la capacidad, c) cálculo de la potencia consumida.

### A. Configuración

Las características del escenario sobre el que vamos a trabajar son las siguientes: se realiza sobre una superficie de  $500 \times 500 \text{ m}^2$ . Las EB están distribuidas de forma aleatoria por el escenario, y los usuarios también, que se mueven de forma aleatoria siguiendo un modelo RWP. Además, se puede diferenciar entre tres tipos de usuarios, dependiendo del número de antenas de transmisión con las que cuenten, siendo dos, tres y cuatro respectivamente los valores que tomarán en el modelo. Por lo que corresponde a las EB, también se realiza una clasificación atendiendo a las características asociadas a dicha EB, pudiendo realizar una división en EB macro, micro o femto.

Las características de cada tipo de EB son mostradas a continuación en la tabla I.

TABLA I  
CARACTERÍSTICAS DE LAS ESTACIONES BASE

	EB Macro	EB Micro	EB Femto
Frecuencia (GHz)	2	5	28
Ancho de Banda (MHz)	10	25	140
Ganancia de antena (dBi)	15	5	5
Transmisión (dBm)	46	25	20

El número de EB y usuarios viene modelado por cada una de las configuraciones existentes. En este caso, se han considerado tres posibles configuraciones donde el número de usuarios es 20, 40 y 50 respectivamente y el número de EB es 5, 14 y 20 respectivamente.

Referente a las conexiones generadas en el escenario, estas siguen un proceso de Poisson con un ratio de llegadas de media  $\lambda = 0,2$ . Donde la duración de una sesión sigue una distribución exponencial con una media  $\mu = 10 \text{ s}$ . En este escenario, las estaciones base están conectadas a una red de acceso donde se encuentran los *routers* que ofrecen conectividad IPv6 a los usuarios móviles con el resto de la red.

### B. Cálculo de la capacidad

Para el cálculo de la capacidad del sistema es necesario primero el cálculo de la relación señal interferencia y para el cálculo de esta es necesario primero calcular la potencia recibida por el usuario, siguiendo con tal propósito la siguiente expresión:

$$P_{RX} [dBm] = P_{TX} [dBm] + G[dB] - L_{PATH LOSS} [dB]$$

Donde  $P_{RX} [dBm]$  es la potencia recibida en dBm,  $P_{TX} [dBm]$  es la potencia de transmisión en dBm,  $[dB]$  es la ganancia de las antenas transmisora y receptoras en dB y  $L_{PATH LOSS} [dB]$  es la pérdida de señal debido al camino de transmisión en dB. A su vez,  $L_{PATH LOSS} [dB]$  se modela de la siguiente manera:

$$L_{PATH LOSS} [dB] = L_{SPACE} [dB] + L_{SHADOW FADING} [dB]$$

Donde  $L_{SPACE} [dB]$  es la pérdida de señal debido a la distancia entre EB y usuario, y que decae siguiendo un exponente de atenuación.  $L_{SHADOW FADING} [dB]$  es la variación en la atenuación debido a múltiples variables como la propagación multirrayecto, cuya distribución sigue una log-normal. Estas pérdidas debido al camino de transmisión se han modelado de seis formas distintas siguiendo tres modelos de transmisión [8], UMi (Urban Microcells), UMA (Urban Macrocells), RMa (Rural Macrocells) para dos casos posibles, LOS (Line-Of-Sight) y NLOS (Non-Line-Of-Sight).

Finalmente, la relación SINR se ha calculado de la siguiente manera:

$$SINR_k = \frac{S}{N} = \frac{P_{rx,j,k}(mW)}{(\sum_{n=1, n \neq j}^M P_{rx,n,k}(mW)) + P_{N_0}(mW)}$$

Donde  $P_{rx,j,k}(mW)$  es la potencia recibida por el usuario  $k$  desde la EB  $j$ ,  $(\sum_{n=1, n \neq j}^M P_{rx,n,k}(mW))$  es la potencia recibida

por el usuario  $k$  desde el resto de estaciones, es decir, la interferencia, y  $P_{N_0}(mW)$  es el ruido debido a las componentes en frecuencia.

Una vez calculada la SINR se calculará la capacidad del sistema como la formula bien conocida para los sistemas MIMO:

$$C_{User,j} = \log_2 \left| I_{NTX_j} + \frac{SNR_{User,j}}{NRX_{User}} * H * H^H \right|$$

Donde  $I_{NTX_j}$  es la matriz identidad con dimensión  $N \times N$ , siendo  $N$  el número de antenas en el receptor, y  $H$  es la matriz del canal. En este caso se ha trabajado con una matriz generada aleatoriamente siguiendo una distribución normal compleja.

### C. Cálculo de la Potencia consumida

Se trabaja con un modelo de consumo de energía que tenga en cuenta tanto el consumo entre el usuario y la estación base como el consumo entre la estación base y el router de acceso. En este caso, el modelo de consumo de energía elegido es el modelo básico de consumo de energía para un transmisor se basa en el modelo presentado en [9- 10]. El consumo regular de energía de una EB, que denotaremos como  $P_{bc}$ , puede expresarse como:

$$P_{bc} = \alpha * P + \beta + \delta * S$$

donde  $P$  representa la potencia transmitida o radiada de cada EB. El coeficiente  $\alpha$  denota la eficiencia de transmisión de potencia debida a un amplificador de RF y a las pérdidas del alimentador, mientras que  $\beta$  representa la potencia disipada debido al procesamiento de la señal,  $\delta$  es una constante que denota un consumo dinámico de energía por unidad de datos, y  $S$  es la velocidad de datos.

La potencia transmitida es la suma de las potencias de los diferentes transmisores más la energía consumida por el backhaul (PBH):

$$P_t = \sum P_{bc} + PBH$$

## III. PLAN DE ASIGNACIONES

Una vez vista la configuración de los parámetros del sistema, así como el cálculo de potencias y SINR dados estos parámetros, se procede a realizar cuatro modalidades de emparejamientos entre los usuarios y las estaciones base, es decir la planificación de las asignaciones de los usuarios a las EBs.

Se presentan de más simple a más compleja, teniendo todas ellas como objetivo maximizar la SINR. Estas asignaciones podrían realizarse siguiendo otros objetivos como la minimización de la distancia o la maximización de la potencia recibida entre usuario y EB. Sin embargo, la maximización de la relación señal-interferencia es la que nos permite obtener mejores resultados en lo que a capacidad del sistema se refiere, proporcionando al usuario mayor velocidades de conexión a la red. A continuación, las cuatro asignaciones.

Planificación 1: se empareja al usuario con aquella EB que proporcione la mayor SINR entre todas las disponibles en el escenario.

Planificación 2: Evita el salto continuo entre EB. Para ello empareja al usuario con aquella EB que proporcione la mayor SINR entre todas las disponibles en el escenario siempre que el cambio provoque una mejora de la SINR por encima de un determinado umbral.

Planificación 3: Se realiza una asignación dual de EB para el mismo usuario. Aquí debe diferenciarse entre EB principal y EB auxiliar, donde la EB principal siempre será una macro celda y la EB auxiliar siempre será una micro o femto celda. De esta forma, la asignación empareja al usuario aquella EB principal que proporcione la mayor SINR entre todas las EB principales y empareja también al usuario aquella EB auxiliar que proporcione la mayor SINR entre todas las EB auxiliares.

Planificación 4: Combinación de la planificación 2 y 3. Toma las ventajas de cada una de ellas evitando saltos continuos de celdas y al ser una asignación dual se mejoran los parámetros del usuario. Se empareja al usuario aquella EB principal que proporcione la mayor SINR entre todas las EB principales siempre que el cambio provoque una mejora de la SINR por encima de un determinado umbral y se empareja al usuario aquella EB auxiliar que proporcione la mayor SINR entre todas las EB auxiliares siempre que el cambio provoque una mejora de la SINR por encima de un determinado umbral.

## IV. OPTIMIZACIÓN DE LOS PARÁMETROS

En esta primera aproximación a la optimización multicapa de un sistema 5G, el parámetro que se va configurar es la potencia de transmisión de las distintas estaciones base, siendo la máxima potencia transmitida la potencia que se encuentra reflejada en la tabla número 1. Para la optimización se ha utilizado la toolbox de Matlab de optimización multiobjetivo donde se han fijado diferentes objetivos.

Primero se hace la optimización buscando 2 objetivos únicamente, la capacidad total del sistema y la potencia consumida. El estudio se realiza con la configuración y optimización 1.

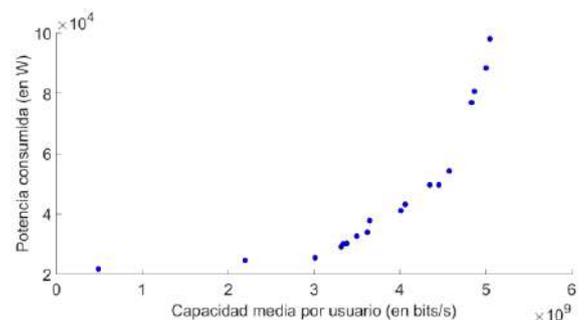


Fig. 1. Optimización conjunta de capacidad por usuario y potencia consumida en el sistema.

En la figura 1 se muestran el frente de Pareto obtenido para esta optimización. El conjunto de soluciones muestra diferentes soluciones (puntos) óptimas, es decir soluciones donde para esa potencia consumida la capacidad es máxima. A la hora de interpretar los resultados, debe tenerse en cuenta que todos los puntos del frente de Pareto son soluciones óptimas, siendo decisión del usuario que punto de operación debe escogerse. En este caso la decisión estará basada en ofrecer un mayor rendimiento en términos de capacidad al

usuario a costa del consumo energético –parte derecha de la figura 1– u ofrecer un menor rendimiento, estando a favor de un mayor ahorro energético –parte izquierda de la figura 1–.

Una forma de contestar objetivamente a la cuestión de qué punto puede considerarse equilibrado es basarse en la eficiencia energética, que muestra el número de bits de información que puede enviarse a través del canal por unidad de energía utilizada. La figura 2 muestra que puntos de operación son los más óptimos en términos de eficiencia energética.

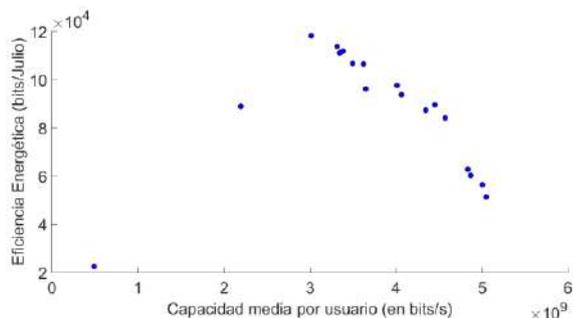


Fig. 2. Eficiencia energética.

Además de los resultados mostrados en la figura 1 y 2, el simulador está preparado para implementar nuevos parámetros clave del sistema gracias a que ha sido implementado tratando de preservar la compatibilidad hacia atrás. A modo de ejemplo de las posibilidades del sistema y su futura ampliación, se muestra un segundo resultado donde se realiza una optimización conjunta de la capacidad y disponibilidad de los usuarios del sistema para la configuración y planificación 1.

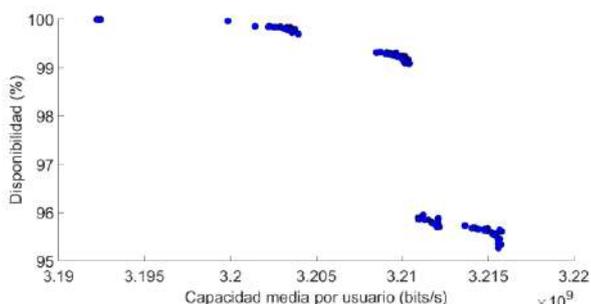


Fig. 3. Optimización conjunta de capacidad por usuario y disponibilidad en el sistema.

Finalmente, se muestra el frente de Pareto donde se optimiza capacidad y potencia obtenido para la configuración 3, que resulta ser más pesada en términos computacionales que la configuración mostrada hasta ahora. Además, se muestra la evolución del frente de Pareto a lo largo de las distintas iteraciones que necesita el algoritmo de optimización para la obtención del frente de Pareto.

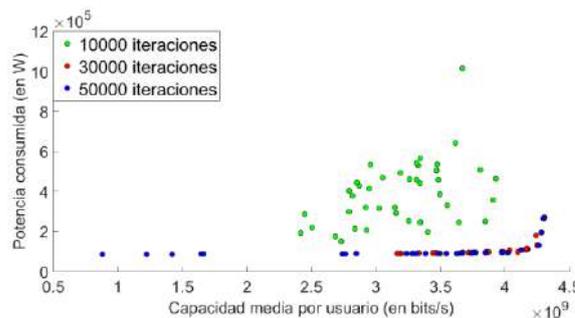


Fig. 4. Optimización conjunta y evolución de capacidad por usuario y potencia consumida en el sistema.

## V. CONCLUSIONES

En este trabajo se ha planteado una optimización de un Sistema 5G con diferentes parámetros de configuración. En estas optimizaciones se muestra como para mejorar un parámetro es necesario perjudicar obligatoriamente a otro de ellos. Una de las métricas más importantes para el futuro de las telecomunicaciones, la eficiencia energética, busca maximizarse a través de estas optimizaciones. Indudablemente, las opciones ofrecidas por el simulador son innumerables, y un trabajo futuro debe desarrollarse para explotar adecuadamente dichas posibilidades.

## AGRADECIMIENTOS

Agradecer a Juanfra y a Pablo todo el apoyo ofrecido, así como la confianza depositada en mí.

Y a mis padres por hacerme ser quien soy a día de hoy.

## REFERENCIAS

- [1] Ericsson Mobility Report June 2018
- [2] Cisco Visual Networking Index: Forecast and Trends, 2017-2022
- [3] E. G. Larsson and L. Van Der Perre, “Massive MIMO for 5G,” *IEEE 5G Tech Focus*, vol. 1, no. 11, March 2017.
- [4] G. Fettweis and E. Zimmermann, “ICT energy consumption: Trends and challenges,” in Proc. WPMC, 2008, pp. 2006–2009.
- [5] M. De Sanctis, E. Cianca, and V. Joshi, “Energy efficient wireless networks towards green communications,” *Wireless Pers. Commun.*, vol. 59, no. 3, pp. 537–552, Feb. 2011.
- [6] Signaling is growing 50% faster than data traffic, Nokia Siemens Network, 2012
- [7] Third-Generation Partnership Project, “Technical specification group radio access network; evolved universal terrestrial radio access (e-utra); potential solutions for energy saving for e-utran (release 10),” Sophia Antipolis, France, Tech. Rep. TR 36.927,V10.0.0, Sep. 2011.
- [8] K. Haneda et al, “5G 3GPP-like Channel Models for Outdoor Urban Microcellular and Macrocellular Environments,” *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, Nanjing, 2016
- [9] F. Richter, A. Fehske, and G. Fettweis, “Energy efficiency aspects of base station deployment strategies for cellular networks,” in Proc. IEEE 70th VTC-Fall, Sep. 2009, pp. 1 -5.
- [10] S. Tombaz et al., “Impact of backhauling power consumption on the deployment of heterogeneous mobile networks,” in Proc. IEEE GLOBECOM, 2011, pp. 1–5.



**Alejandro Ramírez Arroyo**  
(alexrufox@correo.ugr.es) Córdoba, 1997. Graduado en Ingeniería de Tecnologías de Telecomunicación en 2019. Actualmente cursa estudios de máster en Ingeniería de Telecomunicación en la Universidad de Granada interesado en la optimización de redes y el estudio de campos de propagación.

# Diseño y optimización de dispositivos de alta frecuencia para redes 5G

Autor: Miguel Ángel Vaquero Blasco, e-mail: miguelvaquero@correo.ugr.es

Tutores: Juan Francisco Valenzuela Valdés; e-mail: juanvalenzuela@ugr.es

Pablo Padilla de la Torre; e-mail: pablopadilla@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones – Área de TSC  
Universidad de Granada

TABLA I

COMPARATIVA ENTRE 4G Y 5G CON DATOS OBTENIDOS DE [1].

Estándar	Latencia	Throughput	Conexiones	Movilidad
5G	<1ms	10Gbps	1000K/Km <sup>2</sup>	500Km/H
4G (LTE)	30~50ms	100Mbps	10K/Km <sup>2</sup>	350Km/H

**Resumen**—Las redes móviles, al igual que la tecnología y servicios que hacen uso de ella, se han visto en la necesidad de evolucionar y mejorar su infraestructura, aumentando así la calidad y velocidad de los datos que procesan y envían.

Ahora, esta evolución se ha centrado en concebir una nueva generación de redes móviles (5G), dando lugar a un sinfín de propuestas e ideas. Una de ellas se basa en el uso de las técnicas de MIMO (*Multiple Input Multiple Output*) masivo, con lo que se evidencia la necesidad de dispositivos adaptados a este método.

Dentro de este contexto, el propósito del presente proyecto será estudiar y diseñar un modelo de desfasador activo, así como facilitar su inclusión de la manera más optimizada posible en un *array* de antenas, haciendo uso de algoritmos genéticos para tal propósito.

**Palabras clave**—Algoritmos genéticos, *array* de antenas, *Beamforming*, desfasador activo, IOT, microstrip, MIMO, optimización, redes móviles, 5G.

## I. INTRODUCCIÓN

LOS próximos años y, más concretamente, el año 2020, se están convirtiendo en el foco de atención a nivel mundial, y aun hoy la sociedad no es consciente de la dificultad y de la envergadura que supone un proyecto como lo es la evolución de la red móvil tal y como la conocemos.

Como todos sabemos, cada vez es más común que se demanden cantidades ingentes de información, ya sea a través del móvil, tablet, ordenador, o cualquier otro dispositivo conectado a la red. La tendencia de dotar de inteligencia a cualquier objeto que forme parte de nuestra vida cotidiana también es un factor clave en esta evolución.

La llegada del IOT o *Internet of Things*, el avance en vehículos autónomos o la posibilidad de contar con sistemas que trabajen en tiempo real sobrepasan los límites que impone la actual red móvil (4G). Por ende, las demandas que se harán de la quinta generación de redes móviles o 5G exceden las capacidades que en la actualidad se pueden conseguir empleando la red existente, pese a su temprana edad.

Podemos observar una comparativa de las características demandadas en ambos estándares atendiendo a la tabla I. En ella es fácilmente reconocible la superioridad que se espera del 5G frente a las actuales prestaciones de la red.

Resulta evidente pensar que una evolución así precisa de una idea innovadora que genere un proyecto único en su especie. Nada más lejos de la realidad. Un avance de tal envergadura sólo es posible si se aúnan todas las posibles soluciones, encontrando un equilibrio entre todas ellas.

Siguiendo esta filosofía, han surgido distintas soluciones que pretenden solventar las carencias de la red 4G, abriendo así el camino hacia el 2020 y sus objetivos. Algunas de ellas optan por aprovechar de forma más eficiente el espectro de radiofrecuencia, pues esperan que las frecuencias actuales no sean capaces de suplir las demandas del nuevo estándar. Son las denominadas OWC (*Optical Wireless Communication*) y pretender utilizar el espectro de la luz visible, infrarroja y/o ultravioleta.

Pese a no ser la única propuesta que hay, todas ellas se sustentan en tres ideas básicas: rediseñar las infraestructuras actuales destinadas a dar soporte a las comunicaciones para hacerlas más densas y robustas, añadir más elementos radiantes para permitir un mayor flujo de información a través del uso de las técnicas de MIMO, o utilizar un espectro de longitud de onda milimétrica en las comunicaciones. También surgen ideas nacidas como combinación de todas las anteriores, pues una estrecha relación.

Este proyecto se centrará en el estudio de las técnicas de MIMO masivo y, más concretamente, en la forma de optimizar las comunicaciones que empleen estas técnicas. El objetivo de este trabajo radicará pues en proponer un diseño de desfasador activo, optimizado en consumo y rendimiento; así como su inclusión de la forma más eficiente posible en cualquier *array* de antenas. El primero de los objetivos se suplirá haciendo uso de CST Studio Suite, un software de simulación electromagnética que permitirá el trabajo con líneas y circuitos impresos. El segundo de los objetivos se llevará a cabo con la herramienta MatLab, presente en casi cualquier proyecto, a través del uso de algoritmos genéticos, siendo estos los encargados de dar con una solución que optimice tanto consumo como comportamiento del conjunto.

## II. MIMO: PRINCIPIOS DE FUNCIONAMIENTO

Aunque a priori pueda parecer un avance o una técnica moderna, el MIMO está presente en varios aspectos o dispositivos que utilizamos a diario. Así, por ejemplo, cualquier router de Internet que utilice a la vez más de una antena para emitir/recibir información ya es ejemplo de estas técnicas.

La técnica de MIMO, acrónimo de *Multiple Input Multiple Output*, se emplea pues para transmitir o recibir información con sistemas cuya capacidad se ve aumentada al utilizar más de un elemento radiante para tales fines. Un ejemplo sencillo de entender su principio de funcionamiento se puede deducir del sistema de la figura 1.



Fig. 1. Esquema básico de comunicación.

Siguiendo el ejemplo, se cuenta con un emisor/receptor y un canal de comunicación. Sin embargo, es casi imposible encontrar un sistema de comunicación tan sencillo como el mostrado, ya que, en la mayoría de los casos, suelen aparecer obstáculos que imposibilitan la transmisión adecuada de información.

En este contexto aparece el MIMO para dar solución a tal adversidad. Se plantea la posibilidad de contar con más de un elemento radiante en cualquiera de los extremos. Si cada uno de ellos emite la misma información, es sencillo deducir que la probabilidad de recibir la misma sin errores es mucho mayor. Pero eso no dota de mayor capacidad al sistema.

Supongamos ahora un sistema como el que se muestra en la figura 2. En él, aumentamos el número de antenas tanto en emisor como en receptor, por lo que, cualquier información emitida desde alguna de ellas se recibiría en más de una antena, en función de los obstáculos que el entorno presenta.

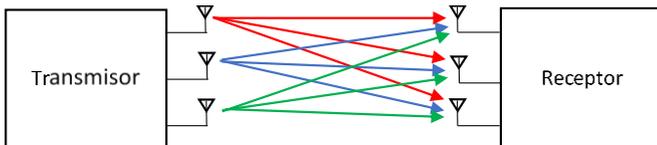


Fig. 2. Ejemplo de comunicación implementando MIMO.

En este punto se presenta la verdadera utilidad de los sistemas MIMO: emitir información distinta con cada una de las antenas. Si aumentamos el número de antenas de forma que aseguremos la integridad de los datos que se envían por cada una de las estaciones emisoras, podemos multiplicar la capacidad del mismo enviando distintas tramas por cada una de ellas. Así, en el sistema de la figura 2 se multiplicaría por tres la capacidad total que se tenía originalmente.

A raíz de este ejemplo se entiende el principio básico del uso de sistemas que implementen MIMO. Lógicamente y, llegados a este punto, es inevitable intentar extender esta técnica pues, a medida que aumentes el número de antenas, la capacidad debería aumentar proporcionalmente y sería a priori sencillo cumplir las demandas del estándar de nueva generación (5G). Es lo que se denomina *massive MIMO* o MIMO masivo.

No obstante, aumentar en exceso el número de antenas puede

ocasionar problemas para transmitir información de forma isotrópica. Concretamente, por cada antena que se incluye en el sistema o conjunto, el patrón de radiación se ve afectado. Si todas las antenas tienen entre ellas un efecto de interferencia constructiva, la señal total emitida se entiende como la suma de todas las componentes o señales que participan en el conjunto. Se obtiene así un sistema de gran capacidad, aunque alteramos su patrón de radiación y el valor de la directividad, si entendemos esta como el cociente entre la potencia radiada en una dirección concreta entre la potencia que se radiaría de forma isotrópica. Podemos apreciar el patrón de radiación de dos sistemas (de 10 y 20 antenas) en la figura 3.

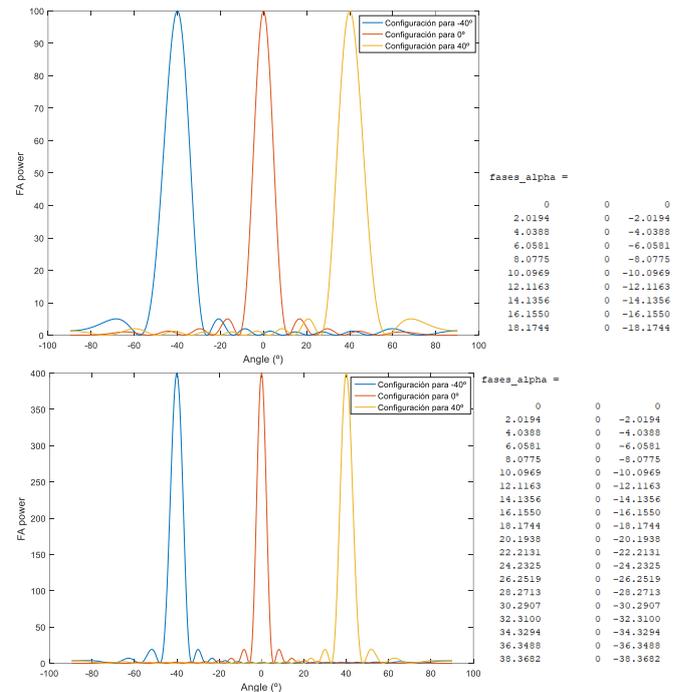


Fig. 3. Configuración y campo generado con 10 (arriba) y 20 (abajo) antenas para los ángulos -40, 0 y 40 grados.

Por último, es imprescindible señalar el carácter “activo” de los desfases que se propondrán en este proyecto, ya que, alterando la fase de las señales con las que alimentamos a cada antena, podemos conseguir que la interferencia sea constructiva en otra dirección (Fig. 3), pudiendo modificar así la orientación de la antena en tiempo real o incluso sin tener que estar presentes, algo que será clave en futuros sistemas que utilicen estas técnicas.

## III. ALGORITMOS GENÉTICOS Y BEAMFORMING

De nada sirve aumentar el número de antenas si el consumo de una sola estación base se dispara. Tenemos que considerar que, por cada antena que se emplee en el conjunto necesitamos un desfaseador para poder modificar la orientación del conjunto a voluntad. No obstante, si analizamos cada caso, las fases de cada antena y su estructura es posible encontrar ciertos patrones y aprovecharse de ellos.

En la segunda parte de este proyecto se buscará dar con la mejor combinación, independientemente del *array* y su estructura, que permita conseguir un consumo óptimo sin renunciar a las prestaciones del conjunto. Esto se realizará implementando una optimización que haga uso de MatLab y sus algoritmos genéticos.

Los algoritmos genéticos tienen un esquema y un funcionamiento que no dista en demasía de los algoritmos convencionales. Por lo general, un algoritmo se concibe como un conjunto ordenado y finito de operaciones que permite hallar la solución de un problema. Aunque son útiles, es difícil optimizar un problema haciendo uso de algoritmos, a menos que usemos métodos de aproximación numérica, como el de Newton-Raphson.

Los algoritmos genéticos, en cambio, permiten resolver cualquier problema cuyas variables a optimizar se puedan cuantificar y expresarse como ecuación. Estos algoritmos repiten el mismo bucle de instrucciones, como haría un algoritmo convencional, con la salvedad de que, en cada iteración, modifican las variables a optimizar de forma arbitraria. Así, cada iteración produce una expresión distinta, y el algoritmo genético descarta las que no minimicen la misma.

Se concibe pues como un sistema similar a la genética de los organismos vivos, basado en la selección natural, donde se producen cambios arbitrarios que hacen desaparecer o perdurar los distintos resultados hasta dar con el más óptimo de todos ellos. Un principio básico de funcionamiento puede verse en la figura 4.

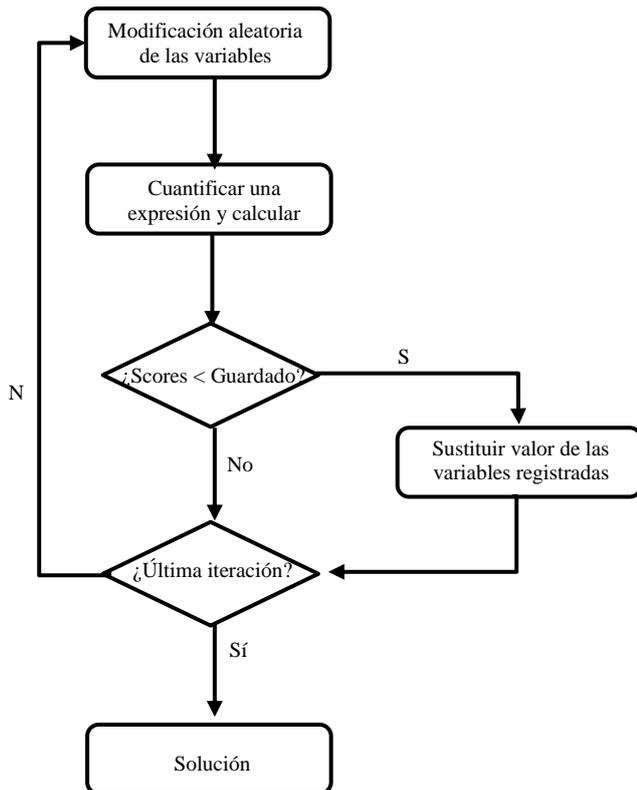


Fig. 4. Base de ejecución de un algoritmo genético.

En este proyecto se pretende aplicar algoritmos genéticos para optimizar el rendimiento de una antena o conjunto de ellas, lo que se conoce como *Beamforming*. Las técnicas de *Beamforming* tienen por objeto el estudio de una configuración de antenas y la búsqueda de una configuración que mejore los resultados, aprovechándose de los efectos de interferencia constructiva de las señales. Un avance en estas técnicas se basa en intentar, no sólo encontrar una solución óptima al problema, sino optimizar también en recursos y, por tanto, en gastos.

En este contexto, se entiende la gran acogida que presenta este proyecto, de cara a optimizar el número de desfasadores activos empleados en cualquier *array*, ya sea unidimensional o bidimensional. Se pretende unificar todos los aspectos a tener en cuenta en una o varias expresiones matemáticas, a fin de minimizarlas o encontrar un equilibrio entre la minimización de todas ellas.

#### IV. DISEÑO DE DESFASADOR ACTIVO

Este apartado corresponde a la primera parte del trabajo realizado, es decir, hará referencia al diseño de desfasador activo propuesto. La propuesta constará de dos partes diferenciadas: la primera, un acoplador híbrido cuya utilidad será dividir la señal entrante para que pueda producirse el desfase, y luego combinar de nuevo las señales desfasadas en el puerto de salida; y la segunda, que se compone de un circuito LC encargado de dotar de una fase deseada a la señal que lo atraviese.

Cabe recordad que ambas partes, si bien forman el desfasador completo, se han realizado por separado y se tratarán en esta memoria de la misma forma. Además, se han realizado dos modelos de desfasadores, centrados al principio y al final de la banda *Ku*, es decir, a 12 y 18 GHz, partiendo de un modelo más limitado que consta de 3 etapas [2].

##### A. Acoplador híbrido

Un acoplador híbrido es un componente que consta de cuatro puertos cuya única función es dividir la señal, a partir de un puerto de entrada, en dos señales de modulo la mitad, que se reparten entre los dos puertos opuestos al de entrada, con un desfase de 90 grados entre ambas.

A priori no goza de mayor utilidad, pero, si hacemos que las dos señales de salida originales sean reflejadas y vuelvan al acoplador, conseguimos que ambas señales se sumen y salgan por el último de los cuatro puertos, que hasta ahora carecía de utilidad.

Este efecto de división, desfase y suma de señales es posible dadas las dimensiones del acoplador. Tanto el ancho como el largo del circuito miden un cuarto de la longitud de onda. Dadas estas características, se producen diversos efectos de interferencia destructiva o constructiva, consiguiendo así que toda la señal que entra por el puerto de entrada se transmita, evitando reflexiones a la entrada. Podemos observar el modelo de acoplador híbrido (Fig. 5), así como las ecuaciones que describen su comportamiento (ecs. 1 a 4).

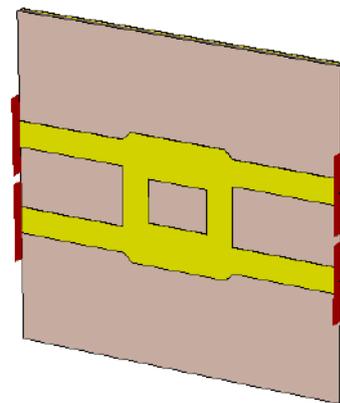


Fig. 5. Diseño de acoplador híbrido centrado en 12 GHz.

$$Puerto\ 2 = \frac{a}{\sqrt{2}} \cdot e^{-j \cdot \omega \cdot \lambda/4} = \frac{a}{\sqrt{2}} \cdot e^{-j \cdot \pi/2} \quad (1)$$

$$Puerto\ 3 = \frac{a}{\sqrt{2}} \cdot e^{-j \cdot \omega \cdot 2 \cdot \lambda/4} = \frac{a}{\sqrt{2}} \cdot e^{-j \cdot \pi} \quad (2)$$

$$Puerto\ 1 = \left( \frac{P_2}{\sqrt{2}} \cdot e^{-j \cdot \pi/2} + \frac{P_3}{\sqrt{2}} \cdot e^{-j \cdot \pi} \right) \cdot e^{-j \cdot \alpha} = 0 \quad (3)$$

$$P_4 = \left( \frac{P_2}{\sqrt{2}} \cdot e^{-j \cdot \pi} + \frac{P_3}{\sqrt{2}} \cdot e^{-j \cdot \pi/2} \right) \cdot e^{-j \cdot \alpha} = a \cdot e^{j \cdot \pi/2} \cdot e^{j \cdot \alpha} \quad (4)$$

Este fenómeno de interferencia entre las señales ocurre en dos etapas. En una primera fase, la señal introducida en el puerto de entrada, o puerto 1, se divide y llegaría a los puertos opuestos al de entrada (ecs. 1 y 2). Al ser reflejada dichas señales, estas volverían a dividirse para salir por los respectivos puertos opuestos. Sin embargo, en el puerto de entrada el sumatorio de las señales es nulo, por lo que toda la señal se transmite por el puerto de salida o puerto 4 (ecs. 3 y 4).

**B. Circuito LC**

Un circuito LC es aquel formado por una capacidad y una inductancia. El efecto conjunto de ambos, aunque dependiente de la frecuencia, implica desfasear una señal que lo atraviese. En este contexto, resultan de gran utilidad para el propósito del desfaseador.

Utilizando la tecnología microstrip, el circuito queda reducido a una línea conectada a una capacidad y a tierra. Para simular el efecto de una inductancia, solo es necesario reducir el grosor de la línea, que, dependiendo de la longitud de la sección que se reduzca, equivaldrá a una bobina de diferente valor. Para dotar al desfaseador de su carácter activo se recurre al uso de un varactor, siendo este un componente que, en función de la tensión existente entre sus bornes, se comportará como una capacidad de valor variable.

A priori, el circuito quedaría resuelto en dos sencillos pasos, aunque el rango de desfases completo que se consiguió en un primer intento fue de 120 grados. Para aumentar dicho rango se hicieron una serie de mejoras: se rediseñó el contacto con tierra, modificando el radio de la corona de la misma hasta rozar el milímetro. A su vez, se modificó la impedancia de la línea al existir relación entre ambos factores. El circuito final se muestra en la figura 6.

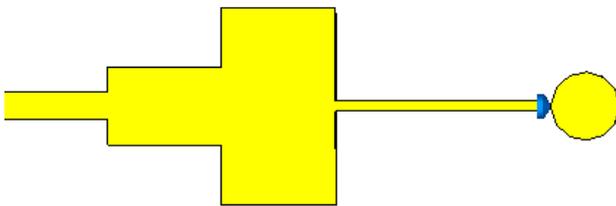


Fig. 6. Diseño final del circuito LC del desfaseador.

**C. Diseños completos a 12 y 18 GHz**

El diseño completo centrado al comienzo y final de la banda Ku, es decir, a 12 GHz y 18 GHz, respectivamente, se muestran a continuación, junto a los parámetros S que los caracterizan. De estos parámetros cabe destacar que expresan la relación entre la señal que entra por un puerto y la que sale por otro de cualquier circuito, expresado normalmente en decibelios. Esto

significa que estos parámetros nos permiten caracterizar cualquier circuito sin la necesidad de conocerlo en detalle.

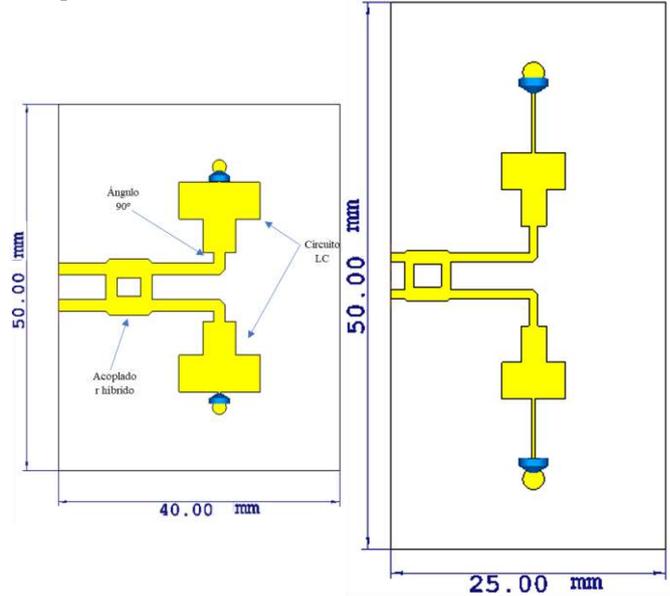


Fig. 7. Diseños completos centrado en 12 GHz (izda.) y a 18 GHz (dcha.).

Como es lógico, en este diseño contemplamos únicamente dos puertos, uno de entrada y uno de salida, por lo que los parámetros S<sub>1,1</sub> y S<sub>1,2</sub> determinan los coeficientes de reflexión y de transmisión del desfaseador, respectivamente. El parámetro S<sub>1,2</sub> expresado en fase, por el contrario, determina el comportamiento del desfaseador en función del valor de capacidad utilizado, y, en definitiva, nos indica la capacidad de desfasear una señal. En cualquiera de los casos, atendiendo únicamente a este último parámetro, es evidente que la señal se transmite casi en su totalidad, albergando un rango de desfases cercano a los 340 grados en el caso de trabajar a 12 GHz; y de 330 grados, si lo hacemos a 18 GHz. Dichos resultados se muestran en las figuras 8, 9, 10 y 11.

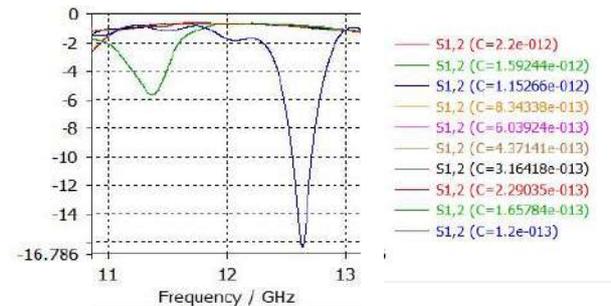


Fig. 8. Parámetro S<sub>1,2</sub>, en magnitud, del desfaseador centrado en 12 GHz.

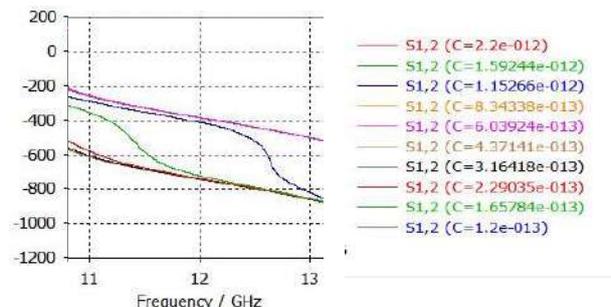


Fig. 9. Parámetro S<sub>1,2</sub>, en fase, del desfaseador centrado en 12 GHz.

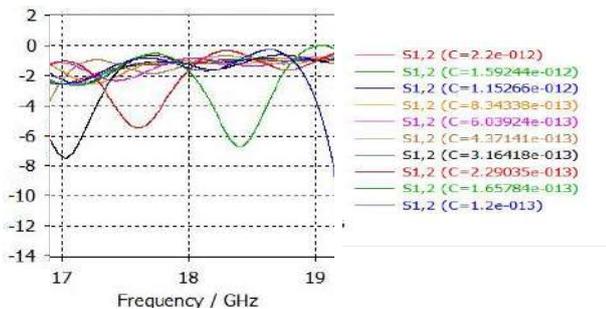


Fig. 10. Parámetro S1,2, en magnitud, del desfaseador centrado en 18 GHz.

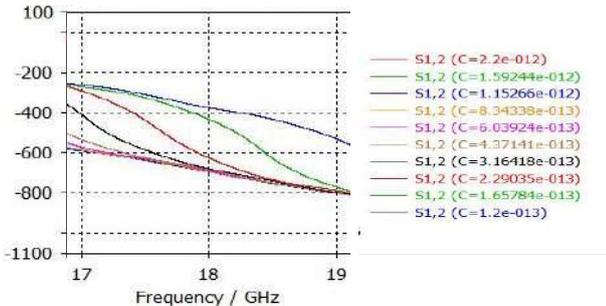


Fig. 11. Parámetro S1,2, en fase, del desfaseador centrado en 18 GHz.

## V. OPTIMIZACIÓN DEL CONJUNTO DE ANTENAS

La segunda parte del proyecto realizado, como ya se ha introducido anteriormente, se centra en encontrar una configuración en cualquier *array* que permita reducir el número de desfaseadores implicados en la misma, pero sin renunciar a la capacidad de orientar el conjunto y obtener el máximo rendimiento del mismo en cualquier dirección.

Se usará para tal propósito la herramienta MatLab, que nos facilitará la labor de realizar un programa o *Script* que haga uso de los ya citados algoritmos genéticos. Diferenciaremos dos programas distintos, uno centrado en la optimización de *arrays* unidimensionales, es decir, donde las antenas se distribuyen a lo largo de una recta; y el segundo adaptado para trabajar con *arrays* bidimensionales, siendo aquellos donde las antenas se distribuyen en un plano.

Pese a contar con dos implementaciones distintas, su principio de funcionamiento es el mismo. En primer lugar, se calculan matemáticamente las configuraciones del *array* en todas las direcciones donde nos interese que funcione adecuadamente el mismo. Partiendo del diseño teórico, realizamos la llamada al algoritmo genético, el cual intentará recrear los diagramas de radiación teóricos, modificando la matriz con las fases de cada antena (de forma arbitraria).

El algoritmo genético tiene por objeto minimizar la denominada “función *Scores*”, siendo esta la cuantización del problema que queremos minimizar. En este caso concreto, se recurre a los algoritmos genéticos multiobjetivo, ya que nos permite elaborar más de una función *Scores* y nos proporcionará diversas soluciones. Estas soluciones se podrán expresar a través de un frente de Pareto, que nos permitirá ver la evolución de las mismas. Como cabe esperar, según la expresión que se minimice, el resultado se acercará mucho o se alejará de la misma forma del caso teórico que se pretende replicar.

Por último, y antes de proceder a mostrar los resultados con ambas configuraciones, es preciso detallar las funciones *Scores* empleadas. Constan de dos etapas. En la primera etapa, el único objetivo radica en conseguir que todos los casos de

estudio tengan un error por debajo del 40 por ciento. Una vez superado ese umbral, el objetivo se bifurca, intentando minimizar por un lado el error obtenido en los casos de estudio, y por el otro, el número de desfaseadores implicados. A esta última función se añade un elemento que nos permitirá apagar la antena en caso de que el efecto que produce sea contraproducente (diodopin).

### A. Resultados obtenidos en arrays unidimensionales

En este subapartado se mostrarán algunos de los resultados obtenidos en *arrays* unidimensionales. Concretamente, se analizarán dos casos, ambos optimizados para operar en tres direcciones, con 10 y 40 antenas, respectivamente. Pueden verse en las figuras 12 y 13.

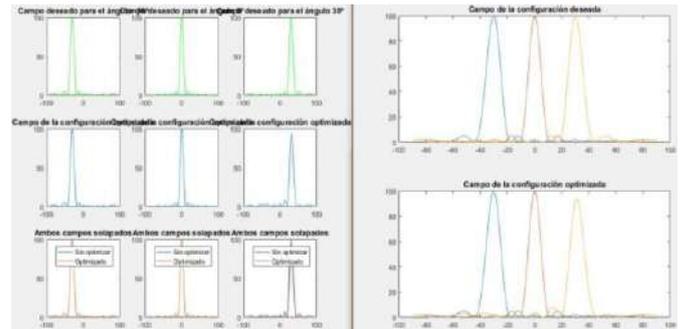


Fig. 12. Optimización realizada en un array de 10 antenas para los ángulos -30, 0 y 30 grados.

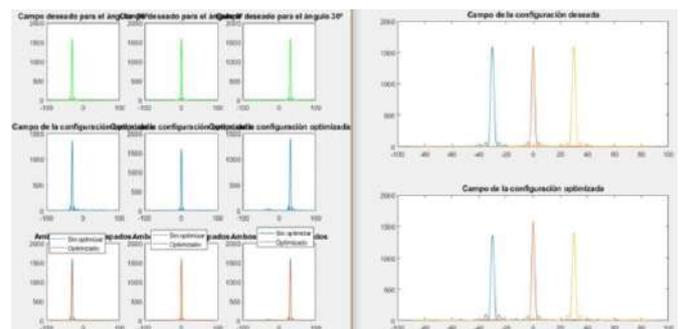


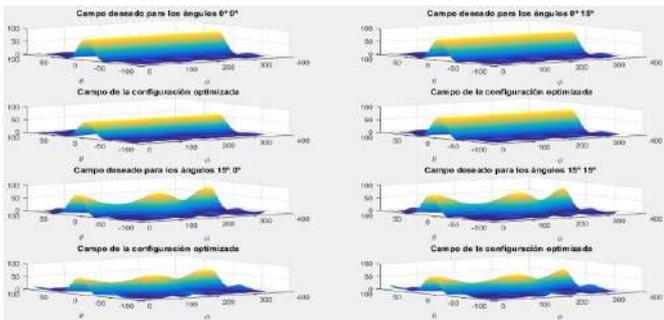
Fig. 12. Optimización realizada en un array de 40 antenas para los ángulos -30, 0 y 30 grados.

Aunque resulta más que evidente, es inevitable resaltar el hecho de que, aun reduciendo en un 30 por ciento el número total de desfaseadores, el error que se percibiría en un sistema real al que se le aplicase la configuración resultante sería ínfimo y no afectaría a las prestaciones del sistema. Es decir, el sistema no vería afectada su capacidad pese a que conseguimos un ahorro, tanto en consumo como en materiales.

### B. Resultados obtenidos en arrays bidimensionales

Por último, se mostrará uno de los resultados obtenidos en la optimización con *arrays* bidimensionales. El hecho de no contar con un número tan elevado de antenas como en el análogo unidimensional se debe en gran parte al elevado cómputo que cada optimización requería. Al tener ángulos y configuraciones bidimensionales, cada iteración del algoritmo se traducía en anidaciones de 4 o incluso más bucles *for*.

Este hecho provocaba un aumento de la carga de trabajo, pero no por ello los resultados obtenidos se han visto afectados, siendo igual de notable la reducción de desfaseadores implicados. Puede comprobarse esta afirmación atendiendo a la figura 13.



Se han empleado 5 de 9 desfasadores y 9 de 9 diodopines en la solución 1. Se han empleado 4 de 9 desfasadores y 9 de 9 diodopines en la solución 2. Se han empleado 4 de 9 desfasadores y 9 de 9 diodopines en la solución 3. Se han empleado 4 de 9 desfasadores y 9 de 9 diodopines en la solución 4. Se han empleado 5 de 9 desfasadores y 9 de 9 diodopines en la solución 5. Fig. 13. Optimización realizada en un array de 3x3 antenas para los ángulos 0,0 y 0,15 grados.

Además de una de las soluciones, podemos observar cómo el algoritmo proporciona un total de 5 soluciones, reduciendo en todas ellas el número de desfasadores en casi o más de un 50 por ciento.

## VI. CONCLUSIONES

Una vez acotado, definido, y realizado el proyecto es primordial realizar una serie de conclusiones, teniendo siempre presentes los resultados obtenidos. A su vez, se tratará de mostrar de forma breve las posibles vías de trabajo futuras que un proyecto como este puede presentar. Cabe mencionar que parte del fruto de esta investigación se ha conseguido plasmar en un artículo [3].

Lo primero que merece mención es que se han cumplido todos y cada uno de los objetivos que se plantearon al comienzo de esta memoria. Visto de otra manera, se ha conseguido obtener un desfasador que, eliminando dos de las tres etapas con las que contaba inicialmente, sea capaz de introducir un desfase, variable a voluntad, de más de 300 grados a una señal que lo atraviese. Además, su inclusión de forma óptima en cualquier *array*, unidimensional o bidimensional, permite un ahorro de consumo y fabricación que aumentan a medida que el conjunto radiante adquiere más elementos.

Pese a los resultados obtenidos y a la resolución satisfactoria del proyecto planteado, no se debe olvidar un aspecto crucial, que habría de estar presente en cualquier proyecto que involucre a la tecnología y el avance de la misma: todo ha de ser susceptible de revisión. Este proyecto no ha de ser un objetivo a corto plazo, como es un Trabajo de Fin de Grado. Nada más lejos de la realidad. Ha de convertirse en una idea en la que seguir ahondando y que en los años venideros pueda ser una realidad con la que seguir experimentando.

Siguiendo esta filosofía, es evidente que el trabajo se ha orientado hacia las futuras redes de 5G, esperando formar parte de ellas. Por ende, resulta lógico pensar que una vía de futuro ha de ser su desarrollo y crecimiento dentro de las mismas. Sin embargo, no debemos olvidarnos de seguir investigando las ideas expuestas, pudiendo constituir otro camino o futuro laboral. Por último, temas como el IOT, la optimización de cualquier ámbito tecnológico, o el avance en transmisión de datos sin fines de formar parte de la red móvil contemplaría otra vía de desarrollo aún por perfeccionar.

## AGRADECIMIENTOS

En primer lugar, me gustaría agradecer a mis tutores, Juan Francisco Valenzuela Valdés y Pablo Padilla de la Torre, su implicación, disposición y ayuda a lo largo de todo el camino que ha supuesto la realización de este trabajo.

También a mi familia, por brindarme la oportunidad de seguir el camino que deseaba sin importar la dificultad del mismo, y por animarme y motivarme tanto personal como académicamente.

Por último, aunque no por ello menos importante, a mis compañeros y amigos, esa segunda familia con la que he compartido cuatro años.

A todos ellos, gracias.

## REFERENCIAS

- [1] Anonymous "NGMN 5G White Paper V1 0,".
- [2] P. Padilla et al, "Electronically Reconfigurable Reflective Phase Shifter for Circularly Polarized Reflectarray Systems," *Lmwc*, vol. 26, (9), pp. 705-707, 2016. Available: <https://ieeexplore.ieee.org/document/7548297>. DOI: 10.1109/LMWC.2016.2597208.
- [3] M. T. ElKhorassani et al, "Electronically tunable phase shifter with enhanced phase behaviour at Ku Band,".
- [4] M. Z. Chowdhury et al, "A Comparative Survey of Optical Wireless Technologies: Architectures and Applications," *Access*, vol. 6, pp. 9819-9840, 2018. Available: <https://ieeexplore.ieee.org/document/8259465>. DOI: 10.1109/ACCESS.2018.2792419.
- [5] R. C. Johnson, *Antenna Engineering Handbook*, 3rd ed. New York, NY, USA: McGraw-Hill, 1996.
- [6] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [7] R. Ouali et al, "Design of electrically tunable phase shifter for antenna arrays perating in ku-band," in 2015, Available: <https://ieeexplore.ieee.org/document/7375467>. DOI: 10.1109/MMS.2015.7375467.



**Miguel Ángel Vaquero Blasco**, 2 de abril de 1996. Graduado en Ingeniería de Tecnologías de la Telecomunicación en la universidad de Granada, con mención en Telemática. Alumno del Master de Desarrollo de Software de la Universidad de Granada.

# High frequency components design for 5G communications

Autor: Ángel Palomares Caballero, e-mail: [angelpc@ugr.es](mailto:angelpc@ugr.es)  
Tutor: Juan Fco. Valenzuela Valdés, e-mail: [juanvalenzuela@ugr.es](mailto:juanvalenzuela@ugr.es)  
Tutor: Pablo Padilla de la Torre, e-mail: [pablopadilla@ugr.es](mailto:pablopadilla@ugr.es)  
Titulación: Máster Universitario en Ingeniería de Telecomunicación  
**Departamento de Teoría de la Señal, Telemática y Comunicaciones**  
**Universidad de Granada**

**Resumen**—A 16-element V-band aperture antenna array based on gap waveguide with glide-symmetric holey structures for future 5G communications has been designed in this M.Sc. Project. With this novel gap waveguide technology combined with E-plane split-block, the power losses and fabrication complexity and costs are fairly decreased. The antenna array consists of two layers that form a feeding network, and several vertical stacked layers which formed the aperture antenna elements. The waveguide size used is WR10 standard. The antenna array design covers a wide bandwidth, 66.5-71 GHz, with a matching impedance level below -10 dB. Also, the antenna array has few gain variation over the entire bandwidth with a gain over 20 dB and global radiation efficiency greater than 80%. The radiation pattern is a pencil beam, with side-lobes lower than -12.5 dB for all the operating frequencies.

**Palabras clave**—V-band array, gap waveguide technology, glide-symmetric structures, aperture antenna, genetic algorithm optimization.

## I. INTRODUCTION

With the arrival of the new generation of communication systems, 5G, the spotlight for satisfying the great demand of data traffic is in the higher part of the frequency spectrum. Recently, the International Telecommunication Union has presented the available frequency bands for 5G communications, including the 66-76 GHz band [1]. In this frame, one of the major challenges to be addressed is the development of radiofrequency (RF) devices and reconfigurable high bandwidth radiant terminals to enable the operation of future 5G systems at those bands. The mmWave band has been center of attention for researchers due to the advances in waveguide manufacturing technologies, which provide an advantage in the level of losses over other technologies such as microstrip or SIW because there is no need of dielectric material. Part of these new waveguide manufacturing technologies is based on gap waveguide technologies [2]. At mmWaves, as frequency is increased, the losses because of an incomplete metallic contact or flaws in the manufacturing of the waveguides are not negligible. The use of gap waveguide permits not to have the requirement of a completely shielded waveguide or complete metallic contact for the different pieces or layers that form the waveguide. Thus, with the use of this technology the losses are almost negligible and the fabrication costs become lower.

There are a variety of antennas operating in the 60 GHz band, based on the gap waveguide technology [3]-[5]. All of these bi-dimensional antenna arrays are formed by three distinguishable parts: the feeding network, some coupling

cavities, and radiating slots. Each of them is implemented in separated layers which are horizontally stacked. The joining surface between layers is the one that includes some elements to avoid the field leakage, such as bed of nails or grid of holes, whose dimensions are defined according to the operating frequency. As the frequency is increased, the physical dimensions of such elements is reduces. That fact can be critical, as the reduction in size implies possible manufacturing challenges and higher manufacturing costs. Recently, a novel gap waveguide technology based on glide-symmetric holey structures has been developed [6]. A periodic structure possesses glide symmetry if its unit cell is generated with a mirroring and a translation of half of the period [7]-[8]. The main effect of this hole-based configuration is that it introduces a wide stopband between first and second propagation modes. This glide-symmetric technology allows a low cost fabrication and a proper performance for gap waveguides, wider in band than the option of bed of nails.

In this Project, this novel gap waveguide technology with glide-symmetric holes is applied for the design and manufacturing of a WR10 based antenna array, operating in the 66-76 GHz band. The structure of this summary Project is the following one. In section II, it is provided some descriptions and simulations of the unit cell of the glide configuration included in the metallic surface of the gap waveguide, the feeding network and the 2x2 aperture antenna subarray. After that, in section III, the final design is presented and the simulation of the complete antenna is provided and discussed. In section IV, the manufactured prototype is showed and finally, the conclusions are drawn in section V.

## II. ANTENNA STRUCTURE

The antenna design is based on the use of gap waveguides for the feeding network and radiating elements of the antenna. The waveguides are designed to have the gap centered at the center of the broad side of the waveguide. This is known as E-plane split-block waveguide. In this guiding structure, the transmission in the guiding waveguide network towards the antenna is carried out with a smaller amount of losses in comparison with a design based on H-plane split-block, with the gap centered in the short side of the waveguide.

Figure 1 provides the antenna array scheme along with some details. The layer configuration of the proposed antenna is the following one: the first two layers are horizontally

placed and form the power distribution network with an input port and four output ports. The rest of the layers are vertically stacked and include the rest of guiding splitters to form the antenna subarray. For avoiding power leakage in the critical gaps of the design (the ones in H-plane), it is used the glide-symmetric holey configuration at the gap surface. This glide-symmetric holey configuration is implemented between the first two layers, and between the second layer and the vertical stacked layers. Finally, each subarray is ended by four aperture antennas which are fed by each one of the output ports of the power distribution network or feeding network.

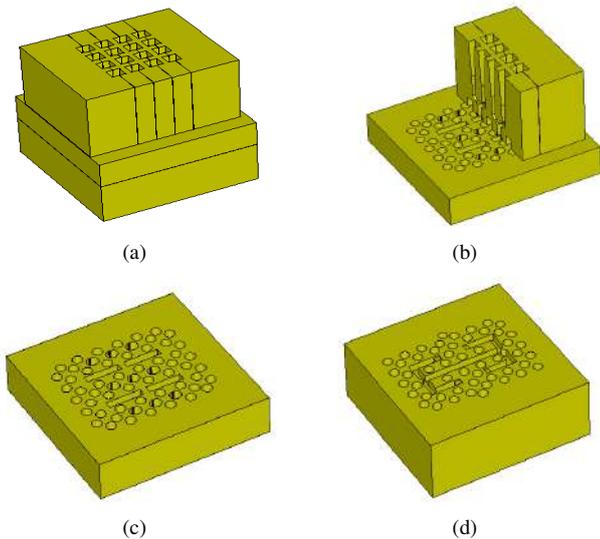


Fig. 1. Antenna array scheme: (a) complete view, (b) separation between the horizontal layers, bottom side (horizontal gap), (c) gap between horizontal layer and rack of vertical layers, (d) cutting view of the complete structure (notice that no glide holes are needed in the vertical gap, since it is located in the center of the broad side of the waveguide)

### A. Unit Cell

A periodic structure possesses glide symmetry if its unit cell is generated with a mirroring and a translation of half of the period. In the case of gap waveguides with holey glide-symmetric structures, the glide-symmetric holes are located over the metallic surface along the guide. Both layers are separated by a very small gap (dozens of microns) which means that there is no perfect shielding and electrical contact between the upper layer and the lower layer. Figure 2 shows the design of the unit cell which forms the holey glide-symmetric structures as well as his dispersion diagram and the stopband that is produced.

The physical dimensions ( $r$ ,  $a$  and  $h$ ) have been chosen avoid the field leakage inside the gap for the desired frequency band, in our case, at least 66 to 76 GHz.

### B. One-to-Four Feeding Network

This section presents the configuration of the initial feeding network. This design is composed of the two horizontal layers shown in Figure 3. The first layer, illustrated in Figure 3(a) and Figure 3(b), contains the input port while the second layer, illustrated in Fig. 3(c) and 3(d), contains the four output ports. Both layers, in their junction surfaces, implement the holey

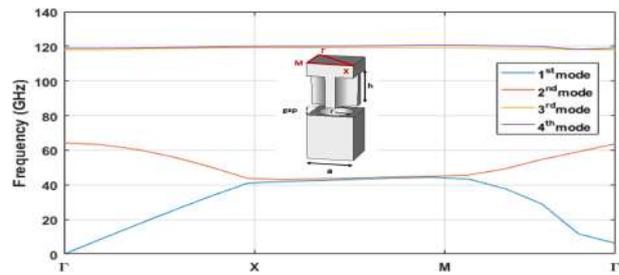


Fig. 2. Dispersion diagram for the glide-symmetric holey structure with the dimensions  $a = 2.26$  mm,  $r = 0.88$  mm,  $h = 2.2$  mm and  $g = 0.03$  mm. Notice the ultrawideband of the bandgap

structures for stopping any leakage produced at the gap. In this design, there is a compulsory gap position change from the center of the waveguide due to a  $90^\circ$  H-plane bend at the end of each output paths of the feeding network. Also, it can be observed that the dimensions of the holes, belonging of each layer, have been chosen also for their use between different paths of the feeding network and thus improve the isolation between them.

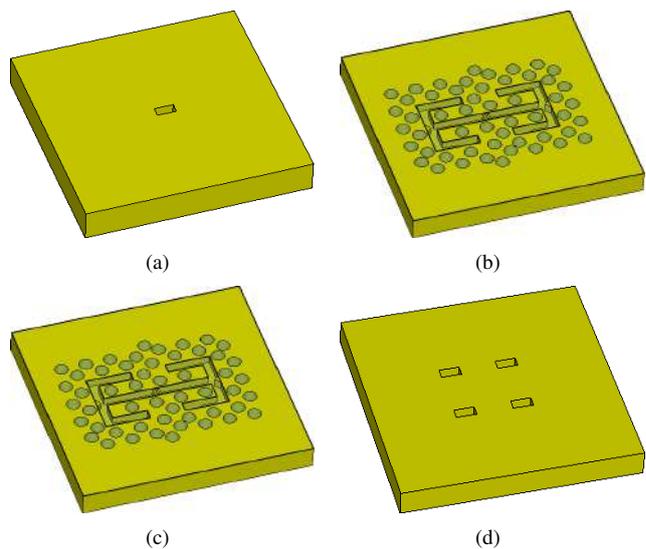


Fig. 3. One-to-Four feeding network Antenna array scheme: (a) detail of the feeding point of the divider (bottom side of the bottom horizontal layer, inverted), (b) bottom side of the middle horizontal layer, (c) top side of the middle horizontal layer (inverted), and (d) the four output ports (top side of the top horizontal layer)

Field simulation of the feeding network design is carried out to calculate the S-parameters and observe the effect of the glide-symmetric holey structures. Figure 4(a) and Figure 4(b) show the reflection and transmission coefficients and the E-field distribution along the structure, respectively.

As can be seen, a good input matching below -15 dB is obtained, and a power transmission to the four output ports of almost -6 dB is reached in a bandwidth from 65 GHz to 71 GHz. In addition, the electric field intensity out of the waveguides that compose the feeding network shows that the field propagation in the gap is blocked and the leakage, avoided.

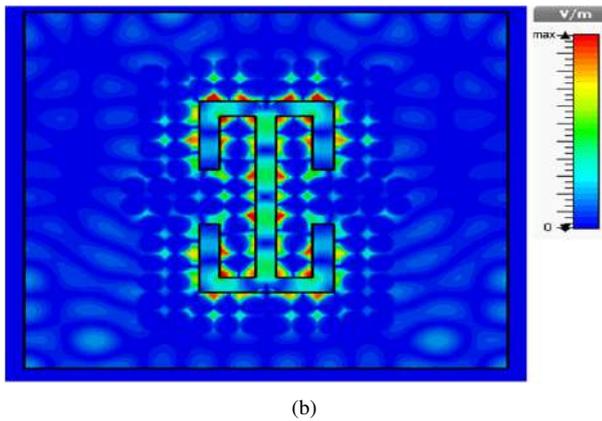
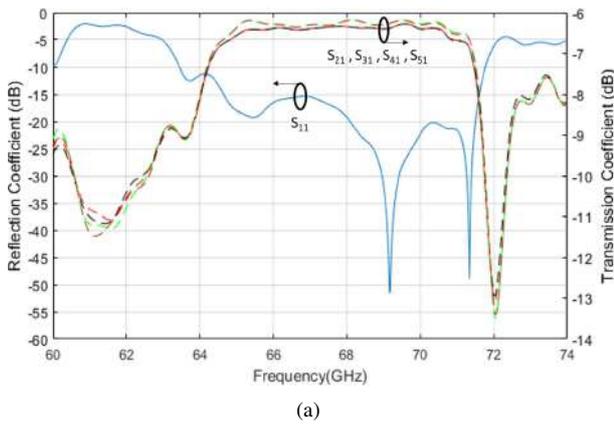


Fig. 4. (a) Simulation results for the reflection and transmission coefficients (amplitude), (b) Simulated electric field distribution of the one-to-four E-shaped power splitter at 69 GHz

C. 2x2 Subarray Antenna

The geometry of the 2x2 subarray, shown in Figure 5(a), is formed by vertical stacked layers and the top of the second horizontal feeding network layer. The vertical stacked layers have in their broad surface the same geometry in order to form the subarray of four aperture antennas. By means of the junction of three vertical layer, it is formed the subarray in which the gap centered in his broad side is also used. The motivation for such E-plane split-block design also in the subarray structure, is for achieve a easy manufacturing without the use of any gap waveguide technology at the both broad surfaces of these vertical layers and for avoid any losses in the signal propagation from the output ports of the feeding network to each aperture antenna. This fact of neglected losses from the output port of the feeding network to the subarray elements can be seen in Figure 5(b) where is shown the electric field distribution in the gap between vertical stacked layers.

The holey structure layer for stop leakage is only necessary in the transition from the output port widening of the upper face of the feeding network, to the start of the waveguides which have been formed by the joint of the vertical layers described before. Thus, the holes are made in both the upper face of the feeding network and the lower part of each vertical stacked layer as it is shown in Figure 6(a) and Figure 5(a) respectively. To check that the effect of the holey structures appears properly, a simulation of electric field distribution inside the gap of the described transition is carried out. The

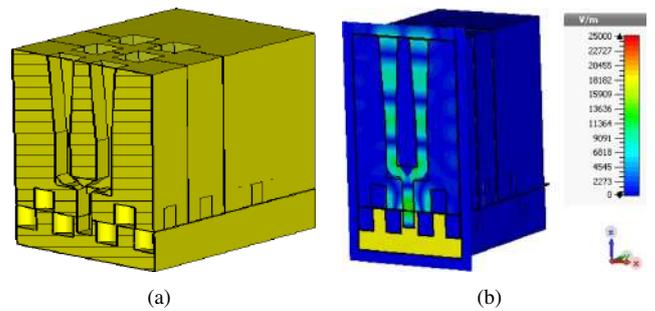


Fig. 5. (a) Geometry, (b) electric field distribution at frequency 69 GHz of the aperture antenna subarray

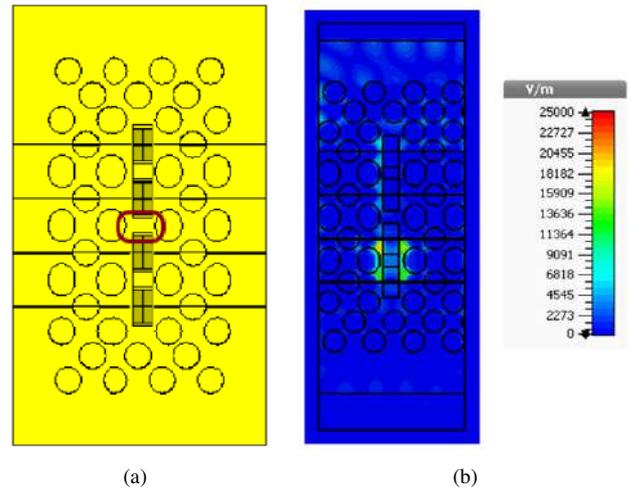


Fig. 6. (a) Geometry, (b) electric field distribution at frequency 69 GHz of the glide symmetry layer

result can be seen in Figure 6(b), where only one port of the two available ports acts as input port.

As it can be observed, there is no leakage and thus power lower losses in the transition. Other measurement that has been conducted is the isolation level between the widening output ports of the feeding network due to there is a side between them where the holey structures can't be used. The isolation level between subarrays structures is performed and the results are shown in Figure 7. It is seen an isolation level below -30 dB in the desired bandwidth and a input matching of the subarray structure below -10 dB in the same bandwidth.

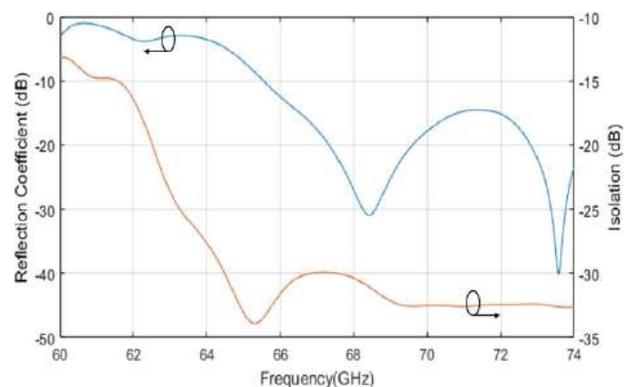


Fig. 7. Simulated reflection and isolation coefficients of the 2x4 element subarray

### III. SIMULATED RESULTS

The complete design of the proposed antenna formed by four 2x2 subarray aperture antennas is illustrated in Figure 8(a). The reflection coefficient is below -10 dB from 66 GHz to 71.8 GHz as is shown in Figure 8(b).

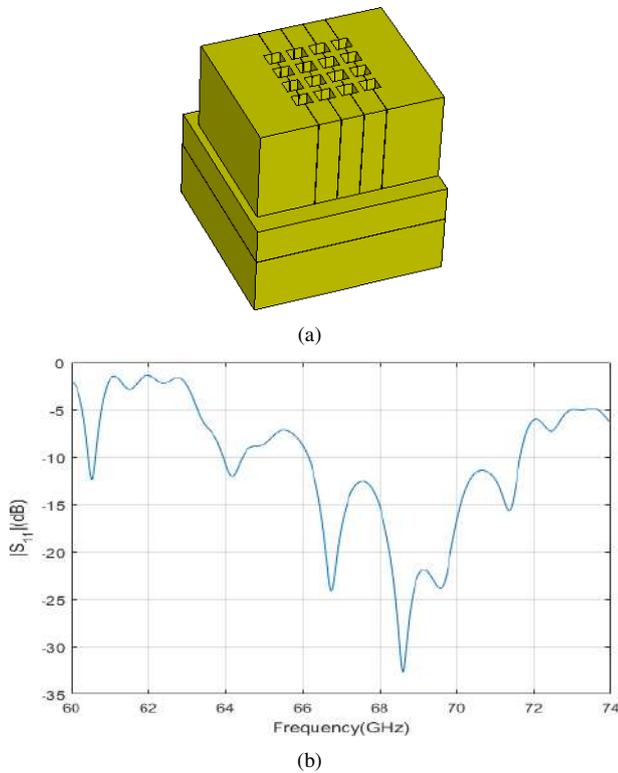


Fig. 8. (a) Complete design, (b) simulated  $|S_{11}|$  of the proposed antenna

Additionally, Figure 9 shows a 3D simulation of the radiation pattern of the proposed antenna. It can be seen that the radiation produced by the antenna is a pencil-beam pattern. Also the gain, directivity and radiation efficiency are carried out by simulation of the proposed antenna in the bandwidth where there is a good input matching. These results are illustrated in Fig. 9.b, where it can be observed that gain and directivity are almost constant from 66.5 GHz to 71 GHz, and greater than 20 dB. Also the radiation efficiency in this bandwidth is bigger than 80%.

Figure 10 shows the normalized radiation patterns of E and H plane for different operating frequencies. As it can be seen, the principal beam direction is  $0^\circ$  and with a good side lobe level, below -12.5 dB, for all the simulated frequencies.

### IV. PROTOTYPE MANUFACTURED

Once the components array design is finished, a prototype is manufactured by SLA (StereoLithography). This 3D printing technique can achieve good resolution both in height and in the XY plane of the manufactured surface, suitable for millimeter-wave electromagnetic components. After the prototype 3D printing, to give a metal bath to the array layers printed for giving them electrical conductivity. The prototype parts and the antenna array assembled are showed in Figure 11. The array layers design is slightly different from the described in previous sections for the need to assemble the array with screws.

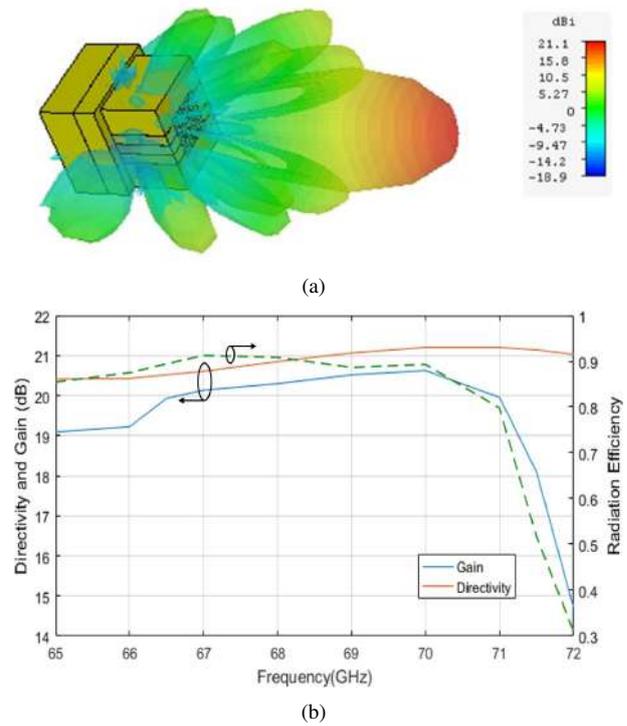


Fig. 9. (a) 3D radiation pattern, (b) simulated gain, directivity and efficiency of the antenna

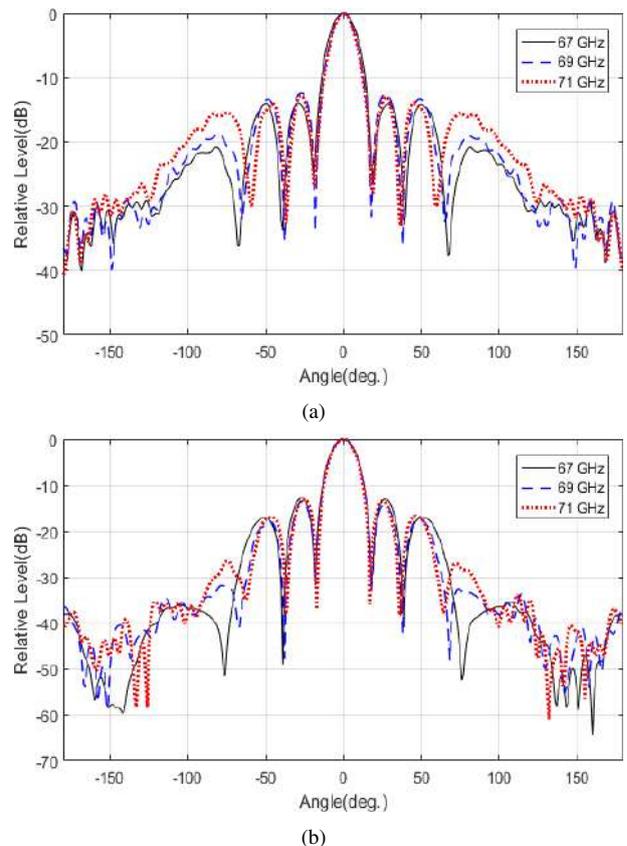
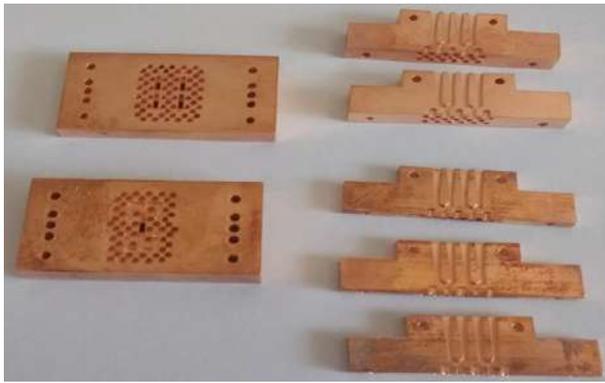
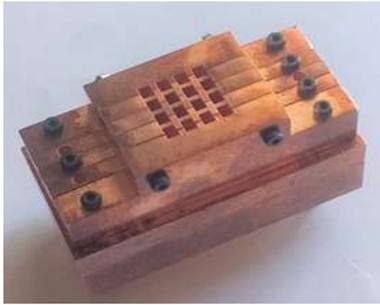


Fig. 10. Simulated radiation patterns of antenna array at frequencies 67, 69 and 71 GHz: (a) E-plane, (b) H- plane



(a)



(b)

Fig. 11. Antenna array prototype manufactured: (a) array layers, (b) array assembled

- [6] M. Ebrahimpouri, O. Quevedo-Teruel and E. Rajo-Iglesias, "Design Guidelines for Gap Waveguide Technology Based on Glide-Symmetric Holey Structures," *IEEE Microwave and Wireless Components Letters*, vol. 27, no. 6, pp. 542-544, June 2017.
- [7] A. Hessel, M.-H. Chen, R.-C. Li, and A.A. Oliner, "Propagation in periodically loaded waveguides with higher symmetries", *Proc. IEEE*, vol. 61, no. 2, pp. 183-195, 1973
- [8] M. Ebrahimpouri, E. Rajo-Iglesias, Z. Sipus, and O. Quevedo-Teruel, "Cost-effective gap waveguide technology based on glide-symmetric holey EBG structures", *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 2, pp. 927-934 Feb. 2018.



**Autor** Ángel Palomares Caballero was born in 1994 in Jaen, Spain. He received his B.Sc and M.Sc. in Telecommunication engineering from the University of Granada (UGR), Spain, in 2016 and 2018, respectively. Since 2017, he has been in the Department of Signal Theory, Telematics and Communications of University of Granada, where he is currently a Ph.D student. His research interests include millimeter wave-antennas, gap waveguide, optimization algorithms and structures with higher symmetries.

## V. CONCLUSIONS

A 16-element aperture antenna array has been designed and simulated. This antenna array has been designed with the use of E-plane split-block and a novel gap waveguide technology based on glide-symmetric holey structures. Simulations results have shown an operating bandwidth from 66.5 GHz to 71 GHz with a reflection coefficient below -10 dB, an almost constant gain greater than 20 dB and a sidelobe level below -12.5 dB in both E and H planes. In addition, fairly good radiation efficiency has been achieved, greater than 80% in all the operating bandwidth. Thus, obtained the performance of the proposed antenna array, it can be concluded that this antenna array is an attractive design for millimeter-wave applications in the future 5G generation as electromagnetic environment interface due to its performance and easy manufacturing.

## REFERENCIAS

- [1] M. J. Marcus, "5G and IMT for 2020 and beyond [Spectrum Policy and Regulatory Issues]," *IEEE Wireless Communications*, vol. 22, no. 4, pp. 2-3, August 2015
- [2] A. Uz Zaman and P.-S. Kildal, "Gap waveguides for mmWave antenna systems and electronic packaging" *Handbook of Antenna Technologies*. Springer, 2016.
- [3] A. Vosough and P. S. Kildal, "Corporate-Fed Planar 60-GHz Slot Array Made of Three Unconnected Metal Layers Using AMC Pin Surface for the Gap Waveguide," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 1935-1938, 2016.
- [4] A. Farahbakhsh, D. Zarifi and A. U. Zaman, "60-GHz Groove Gap Waveguide Based Wideband H-Plane Power Dividers and Transitions: For Use in High-Gain Slot Array Antenna," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 11, pp. 4111-4121, Nov. 2017.
- [5] D. Zarifi, A. Farahbakhsh, A. U. Zaman and P. S. Kildal, "Design and Fabrication of a High-Gain 60-GHz Corrugated Slot Antenna Array With Ridge Gap Waveguide Distribution Layer," *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 7, pp. 2905-2913, July 2016.



# Integración del simulador de canal de propagación *QuaDRiGa* en un simulador de nivel de sistema

Autor: Francisco Jesús Quero de la Rosa, e-mail: fjqr@correo.ugr.es  
Tutor: Pablo Muñoz Luengo, e-mail: pabloml@ugr.es  
Co-tutor: Juan Francisco Valenzuela Valdés, e-mail: juanvalenzuela@ugr.es  
Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—La evolución de las tecnologías inalámbricas está marcada por el gran crecimiento de usuarios de las infraestructuras de comunicación. Con la finalidad de satisfacer los rigurosos requisitos de las nuevas tecnologías de radiocomunicación y facilitar su planificación y despliegue, se han impulsado iniciativas por parte del sector industrial y del sector académico. Una de ellas es *QuaDRiGa*, una herramienta de simulación utilizada por diversos proyectos de investigación para generar respuestas de canales de radio. Este proyecto amplía las funcionalidades de *QuaDRiGa* para producir simulaciones a nivel de red de escenarios 5G, con la adición de ciertas mejoras que hacen que las simulaciones otorguen resultados significativos. Además, se realiza una serie de evaluaciones del simulador en variedad de escenarios para así determinar cómo afectan las configuraciones de la infraestructura al desempeño de la misma en términos de capacidad y cobertura.

**Palabras clave**—5G, Capacidad de Canal, Comunicaciones de Radio, Comunicaciones Inalámbricas, IoT, HetNets, Macro-Celdas, Micro-Celdas, *QuaDRiGa*, Redes 5G, Simulador, Simulador de Nivel de Sistema.

## I. INTRODUCCIÓN

EL despliegue de infraestructuras de nuevas tecnologías resulta una tarea que puede conllevar numerosas labores de estudio y grandes costes, tanto de infraestructura como de mano de obra [1]. Además, debido a la creciente complejidad de las mismas, el desarrollo requiere una mayor atención, especialmente en escenarios heterogéneos (*Heterogeneous Networks*, *HetNets*), Figura 1, como lo es el futuro estándar 5G cuyo punto fundamental es precisamente la coexistencia de distintos tipos de estaciones base (*Base Stations*, *BSs*).

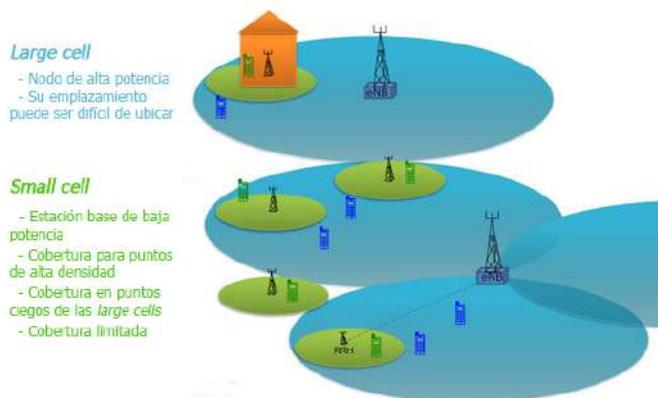


Fig. 1: Entorno típico de *HetNets* [2]

Por tanto, anticiparse a futuros problemas, así como cerciorarse de que el despliegue de las nuevas redes ofrece un servicio correcto y completo, son tareas primordiales a la hora de la planificación del acceso a través de diferentes tecnologías. Tanto para los proveedores de servicio e infraestructura como para los fabricantes de cualquier tipo de sistema 5G, es necesario conocer el comportamiento de sus sistemas previamente a su implantación, con la finalidad de optimizar su funcionamiento y corregir posibles problemas.

Con el objetivo de hacer frente a esta clase de tareas, existen ciertos proyectos con la iniciativa de desarrollar un simulador que sea capaz de realizar los cálculos oportunos para conocer con detalle los datos característicos -pérdidas de propagación, respuesta del canal, niveles de interferencia, etc.- de un entorno ficticio pero realista de tecnología 5G.

Un claro ejemplo es el de *QuaDRiGa* (QUAsi Deterministic RadIo channel GenerAtor) [3], un proyecto que surgió con la financiación del programa europeo FP7, creado por investigadores del *Fraunhofer Institute for Telecommunications* [4]. Aunque actualmente existen alternativas como *NYUSIM* [5], *QuaDRiGa* es especialmente interesante debido a su potencial de uso ya que sus modelos son concebidos para ser compatibles con estándares de entidades como 3GPP [6] o WINNER [7] y está siendo utilizado en actuales proyectos del *5G Infrastructure Public Private Partnership*, *5GPP* como por ejemplo *mmMAGIC* [8].

*QuaDRiGa* es un conjunto de funciones escritas para Matlab/Octave distribuido con licencia de software libre GNU 3.0 que permite generar los coeficientes del canal de un entorno de red a través de modelos de canal tanto de *Long Term Evolution*, *LTE*, como de 5G, permitiendo la simulación de un entorno heterogéneo con multitud de características.

Este trabajo presenta la integración del generador de canal *QuaDRiGa* en un simulador de redes 5G que es capaz de medir el rendimiento de escenarios multi-celda configurables, basándose en *QuaDRiGa* para generar los coeficientes del canal y posteriormente hacer los cálculos oportunos, como por ejemplo el *throughput* o las interferencias recibidas de otras celdas.

Tabla I: Características estándar de las celdas simuladas

	Altura	Dist. entre celdas	Frecuencia	Entorno
Macro-celda	25 m	2000 m	2 GHz	Urbano
Micro-celda	10 m	500 m	60 GHz	Urbano

En concreto, este simulador concibe dos tipos de celda: macro-celda y micro-celda urbanas, cuyas respectivas características predeterminadas se han establecido de acuerdo con las del Reporte Técnico 38.901 de 3GPP [9], como se muestran en la Tabla I. Además, el simulador también permite la elección de parámetros como número de terminales móviles -los cuales se encuentran en movimiento-, número de estaciones base, frecuencias centrales, potencia de transmisión y ancho de banda por celda. A partir de este Trabajo de Fin de Grado ha surgido una publicación en el congreso URSI 2018 la cual recibió una muy buena acogida.

El presente informe se organiza en cuatro secciones además de la actual introducción. La Sección II realiza una introducción a *QuaDRiGa* así como una descripción de las primeras mejoras que se han implementado, como una visualización multi-celda detallada y la convivencia de varios tipos de celda en un mismo escenario, necesarios para una simulación compleja coordinada en todas sus capas. La Sección III describe todo el desarrollo del simulador donde se implementan los cálculos de capacidades, interferencias y emparejamientos. La Sección IV presenta un ejemplo de uso del simulador con sus resultados. Finalmente, en la Sección V se encuentran las conclusiones.

## II. INTEGRANDO *QuaDRiGa* EN 5GNERALIFE

Como se comentaba anteriormente, *QuaDRiGa* es un software desarrollado para Matlab y Octave cuya función es la generación de los coeficientes de canal de radio a partir de una configuración deseada.

Dichos coeficientes son generados a partir de elementos que en *QuaDRiGa* están definidos como clases. Se pueden distinguir los objetos de las principales clases: *Layout* (capas) y *Channel* (canal), mientras que se utilizan objetos de otras clases a modo de auxiliares, por ejemplo *ArrayAnt*, para modelar las antenas de las estaciones base, o *SimParam*, que especifica los parámetros de simulación de cada *Layout* como la frecuencia central.

Aunque a priori pueda parecer complejo, las pautas de utilización de *QuaDRiGa* siempre siguen los mismos pasos. Sin embargo, a la hora de complementar un generador de canal con la finalidad de constituir un simulador de red funcional, surgen diferentes problemas que solventar así como el desarrollo de distintos métodos que permitan cohesionar todos los elementos generados para así obtener el simulador con las funciones deseadas.

Por ello, el simulador implementado hace frente a dichos problemas a través de un uso automatizado de *QuaDRiGa* con la finalidad de generar canales de comunicación de una red con características 5G, de entre las que se encuentra la posibilidad de simular frecuencias de hasta 100 GHz, uso de dos capas de celdas para generar un entorno heterogéneo, características de estaciones base -como altura, radiación, etc.- y dar al usuario la elección de parámetros como la potencia de recepción o ancho de banda.

A continuación, se presentan las pautas que el simulador sigue para la generación de escenarios y simulaciones a partir de *QuaDRiGa*:

- 1) El primer paso es establecer los parámetros de simulación para cada *Layout*, esto es, especificar frecuencia

central entre otros parámetros como el uso de ondas esféricas, o el modelo de polarización, así como otros aspectos avanzados que no son objetivo de estudio de este proyecto ya que *QuaDRiGa* incorpora por defecto una configuración que se ajusta con coherencia a los parámetros de 3GPP.

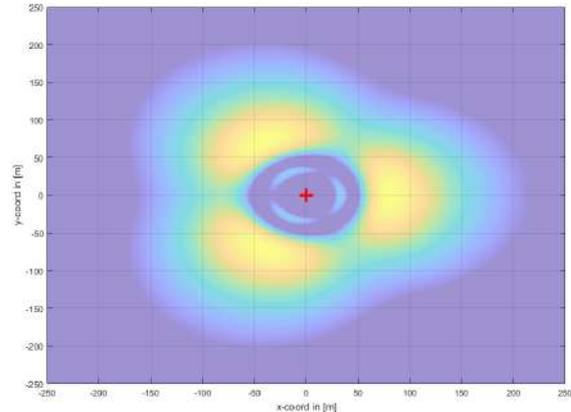


Fig. 2: Esquema de radiación de la antena utilizada para estaciones base

- 2) A continuación, se crean tantos *Layouts* como tipos de celda se desean y/o por cada frecuencia a utilizar. Para generar dichos *Layouts* es necesario asignarles los parámetros del primer paso, así como el número de estaciones base y de terminales móviles. Debido a limitaciones que el propio *QuaDRiGa* presentaba, se decidió modelar ambas celdas utilizando una estación base en el centro de las mismas que utiliza tres antenas sectoriales con una apertura horizontal de  $120^\circ$  de modo que tratan de aproximarse a una antena omnidireccional, o más apropiadamente, al lóbulo que una antena sectorial radiaría desde el vértice de uno de los hexágonos que las celdas formarían, véase la Figura 2. Destacar también que para una simplificación de los cálculos y por más limitaciones de *QuaDRiGa*, especialmente por el hecho de que no permite el uso de más de una frecuencia por *Layout*, se ha implementado un reuso unitario de frecuencias.
- 3) Una de las tareas más importantes que el simulador realiza, puesto que cada capa o *Layout* es independiente entre sí, es forzar la compartición de terminales móviles entre *Layouts*, de modo que en todas estas capas adquieran las mismas características, como su movimiento o posición. Además, se ha implementado el modelado de movimiento urbano de los terminales móviles, así como la probabilidad de que un terminal móvil se encuentre en visión directa con la BS. Además, cada terminal móvil recibe un total de 20 trayectorias de la señal cuando no se encuentra en visión directa, modelado que deriva del propuesto en el 3GPP-SCM [10] como solución para poder calcular la trayectoria combinando interferencias constructivas y destructivas.
- 4) El último paso que implica el uso de *QuaDRiGa* es generar los coeficientes de canal de cada *Layout*. Como se comentaba en el paso 3, estos canales son indepen-

dientes sin tener en cuenta sus terminales móviles ni la frecuencia, por lo que la labor de la construcción del simulador debe centrarse en combinarlos en esta etapa de desarrollo. Para combinar los canales se necesita crear un objeto *Channel* a partir de varios objetos de la misma clase, lo que implica que el objeto resultante contendrá información sobre los coeficientes del canal en su conjunto, combinando así todos los elementos independientes. De este modo, se puede trabajar con interferencias y capacidades en futuros pasos.

### III. DESARROLLO PRINCIPAL DEL SIMULADOR

Una vez que se ha utilizado *QuaDRiGa* -con sus respectivas mejoras- para la obtención de los coeficientes de canal, se puede abordar el procesado de los mismos con el propósito definitivo de obtener datos sobre el rendimiento del escenario heterogéneo simulado.

Esta etapa es la más compleja debido a la dificultad de la concepción de las funcionalidades a implementar en el simulador, ya que *QuaDRiGa* no admite de manera nativa el procesado de los coeficientes de canal generados para obtener datos característicos como capacidad o interferencias, por lo que fue necesario implementarlas desde cero.

El objetivo de esta fase es conseguir, a partir de los coeficientes de canal generados, una o varias medidas del rendimiento del entorno, de modo que se pueda estudiar cuantitativamente el desempeño de la red en función de diversas variables como distintos criterios de emparejamiento terminal-estación base, la distinta naturaleza de las celdas, la variación del ancho de banda o los distintos modelados del movimiento y la posición de los terminales móviles. Para integrar esta funcionalidad, se ha desempeñado una labor de desarrollo que constaba de los siguientes pasos:

- 1) Cálculo de potencia recibida en cada terminal para cada estación base. A partir de los coeficientes obtenidos en anteriores pasos, se realizan los cálculos oportunos para convertir esos coeficientes a potencia recibida, teniendo en cuenta que se encuentran normalizados.

$$P_{Rx}(dBm) = \log_{10}(|Coeffs|^2) + P_{Tx}(dBm)$$

- 2) Cálculo de SINR (*Signal-to-interference-plus-noise ratio*). Una vez que se obtienen los valores de potencia, para cada terminal móvil, se recopila la potencia recibida por cada estación base, ya sea micro- o macro-celda, con el propósito de utilizarla en el cálculo del SINR entre cada estación base y cada terminal:

$$SINR_i(W) = \frac{P_{BS}(i)}{N_o - P_{BS}(i) + \sum_{j=1}^{N_{BS}} P_{BS}(j)}$$

donde  $P_{BS}$  es la potencia recibida de la estación base que se quiere evaluar, en Watios,  $N_{BS}$  el número total de estaciones base y  $N_o$  es el ruido térmico, que por defecto se ha implementado con una densidad espectral de  $4.04e-21$  W/Hz.

- 3) Atendiendo a diversos criterios, se crean emparejamientos atendiendo a diversos criterios. Teniendo todos los datos anteriores, se decide qué estación base sería la más adecuada para crear un enlace con cada uno de los terminales teniendo como criterios independientes

la estación base mas cercana, la de mayor potencia recibida y la de mayor SINR.

- 4) A partir del ancho de banda asignado a cada celda y a los resultados de los emparejamientos anteriores, se calcula la capacidad del canal para cada terminal en cada una de las situaciones:

$$C = BW * \log_2(1 + SINR)$$

donde  $BW$  es el ancho de banda asignado a la estación base en cuestión, en Hz, y  $SINR$  es el valor calculado en anteriores pasos.

Con esta instanciación de parámetros y de criterios, se ha logrado implementar una serie de funciones que permiten observar cómo evoluciona con el tiempo el rendimiento de la red así como el traspaso de una celda a otra.

Además, se ha desarrollado una mejora que permite visualizar todas las capas de simulación al mismo tiempo, con la finalidad de contar con una muestra tangible de la disposición del escenario de simulación, con la adición de las fronteras entre celdas -en rojo las macro-celdas y en verde las micro-celdas- como se podrá observar en la siguiente sección.

### IV. REALIZACIÓN DE SIMULACIÓN DE DEMOSTRACIÓN

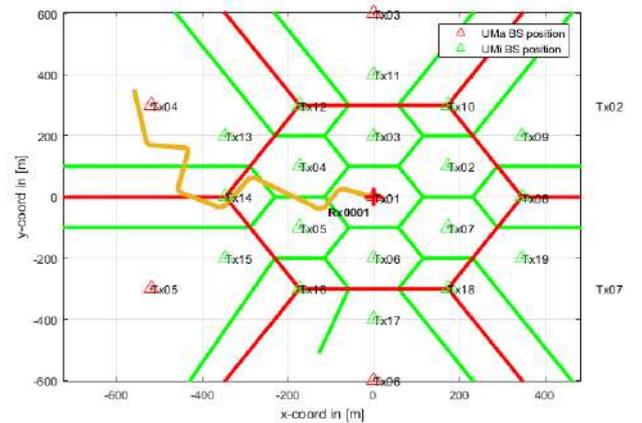


Fig. 3: Escenario de simulación.

Para ilustrar el funcionamiento del simulador, se ha procedido a la simulación de una realización de prueba constituida por únicamente un terminal móvil que se desplaza a lo largo de un recorrido de 1 km en un entorno formado por una capa de micro-celdas de 17 estaciones base y una capa de macro-celdas con siete estaciones base. Además, la potencia de transmisión es de 40 dBm y 27 dBm para macro-celdas y micro-celdas respectivamente, y el ancho de banda asignado a cada estación base es de 100 MHz. De acuerdo con la sección II, el reuso de frecuencias es unitario.

En la Figura 3 se ha representado la trayectoria del receptor y la posición de las correspondientes estaciones base, con la delimitación geométrica de las celdas. El modelo de canal utilizado es el de 3GPP-TR38.901 para macro-celdas y micro-celdas urbanas [9].

Para mostrar datos significativos de la simulación, se ha variado la frecuencia de trabajo a 6 GHz, 20 GHz y 60 GHz

## V. CONCLUSIONES

Se ha presentado el procedimiento de elaboración de un simulador para entornos heterogéneos orientado a 5G y basado en el generador de canal *QuaDRiGa*. Este simulador admite como parámetros variables la frecuencia central de cada tipo de celda, el número de terminales móviles, la distancia recorrida por los portadores de los terminales, el ancho de banda de cada celda.

El simulador ha resuelto algunas de las limitaciones de *QuaDRiGa* al extender el uso de la herramienta al nivel de sistema. No obstante, debido a limitaciones propias de *QuaDRiGa* y de simplificación de implementaciones del simulador, aún quedan aspectos que se podrían mejorar y perfeccionar como el modelado del esquema de radiación de cada celda, el reuso de frecuencias o la incorporación de mayor cantidad de criterios de emparejamiento, incluso el uso de algoritmos de optimización para obtener la localización geográfica óptima de las estaciones base para una determinada configuración.

Con vista a los próximos pasos, se está diseñando actualmente unas indicaciones de uso, para poder simular multitud de escenarios propios de 5G, en los cuales se pueda evaluar las prestaciones de la red para una configuración determinada y diferentes modificaciones de los parámetros, de modo que esto sirva para una correcta planificación y optimización de la red de acceso radio. También se está estudiando la posibilidad de implementar la opción de aprovechar la simulación de entorno heterogéneo para utilizar técnicas como por ejemplo coordinación de multi-punto (*Coordinated Multi Point, CoMP*) para así estudiar el rendimiento en redes que utilicen técnicas avanzadas.

## AGRADECIMIENTOS

Agradecer a mis tutores Pablo y Juan Francisco la labor que durante el curso 2017/18 desempeñaron mentorizando el presente proyecto. Gracias a ellos no solo 5Gneralife vio la luz, sino que aprendí casi todo lo que sé sobre el complejo mundo de las comunicaciones inalámbricas.

## REFERENCIAS

- [1] "Mobile subscriptions worldwide outlook," <https://www.ericsson.com/en/mobility-report/reports/november-2017/mobile-subscriptions-worldwide-outlook>, November 2017.
- [2] J. Wannstrom and K. Mallinson, "Hetnet/small cells," 2014.
- [3] S. Jaeckel, "Quadriga: the next generation channel model," <http://quadriga-channel-model.de/>, 2014.
- [4] F. Burkhardt, S. Jaeckel, E. Eberlein, and R. Prieto-Cerdeira, "Quadriga: A mimo channel model for land mobile satellite," in *The 8th European Conference on Antennas and Propagation (EuCAP 2014)*, April 2014, pp. 1274–1278.
- [5] T. S. Rappaport, S. Sun, and M. Shafi, "5g channel model with improved accuracy and efficiency in mmwave bands," *IEEE 5G Tech Focus*, vol. 1, no. 1, pp. 1–6, 2017.
- [6] "The 3rd generation partnership project (3gpp)," <http://www.3gpp.org/about-3gpp>.
- [7] M. Narandžić, C. Schneider, and R. Thomä, "Winner wideband mimo system-level channel model, comparison with other reference models," in *Proceedings of the Internationales Wissenschaftliches Kolloquium (IWK'09)*, vol. 54, 2009.
- [8] S. Jaeckel, M. Peter, K. Sakaguchi, W. Keusgen, and J. Medbo, "5g channel models in mm-wave frequency bands," in *European Wireless 2016; 22th European Wireless Conference*, May 2016, pp. 1–6.
- [9] G. T. . v14.1.0, "Study on channel model for frequencies from 0.5 to 100 ghz," 2017.
- [10] G. T. . v14.0.0, "Spatial channel model for multiple input multiple output (mimo) simulations," 2017.

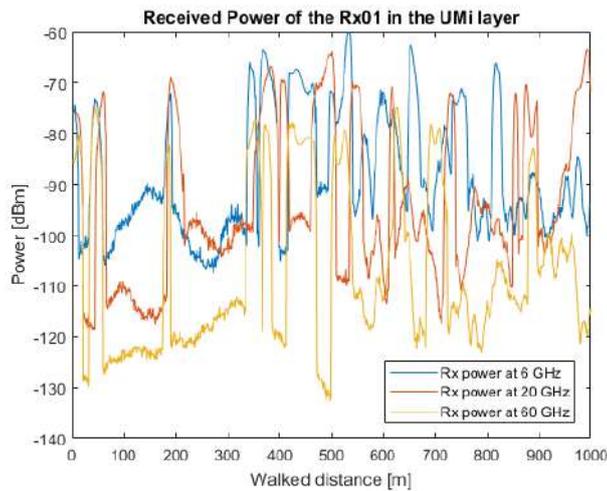


Fig. 4: Comparativa de la potencia de recepción para la capa de micro-celdas a distintas frecuencias.

para la capa de micro-celdas, representando la potencia de recepción para cada uno de los casos, como se muestra en la Figura 4. En ella se puede apreciar el desvanecimiento de la señal, el cual es significativamente mayor en caso de frecuencias altas.

Del mismo modo, si se obtienen los valores de la capacidad asignada según las diferentes estrategias de emparejamiento, mostrados en la Figura 5, que estación base de mayor SINR y estación base más cercana, se observa que aunque se obtienen valores parecidos en ambos casos, el caso de SINR es estrictamente mejor gracias a que mejora las condiciones en ciertos tramos. Cabe destacar que en casos con mayor número de usuarios, la capacidad puede resultar favorable en el caso de emparejamiento por cercanía, teniendo en cuenta que, aunque se reciba mayor SINR de otra estación base, la capacidad se reparte entre todos los terminales, por lo que pueden existir celdas con menor utilización que resulten óptimas.

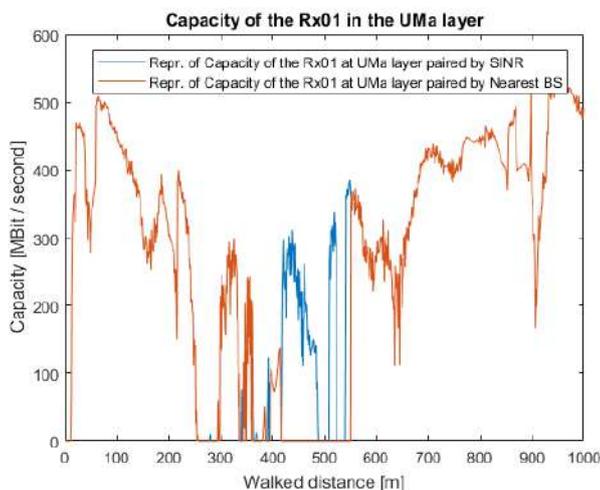


Fig. 5: Comparativa de capacidad máxima asignada para diferentes estrategias de emparejamiento.

# Interconexión de Miembros de un Equipo de Emergencias en Entornos sin Cobertura usando Dispositivos Móviles

Autor: Félix Delgado Ferro, e-mail: felixdelgado@correo.ugr.es

Tutor: Jorge Navarro Ortiz, e-mail: jorgenavarro@ugr.es

Titulación: Grado en Ingeniería en Tecnologías de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—Hoy en día, las tecnologías referentes al Internet de las Cosas incrementan sus integraciones exponencialmente en la sociedad actual. Estas tecnologías suelen emplear elementos de bajo consumo que se emplean para la creación de redes de amplio alcance a bajo coste.

En este trabajo se describe una red diseñada para su funcionamiento en zonas sin cobertura empleando tecnologías diversas para las comunicaciones como Bluetooth Low Energy y Long Range Wide Area Network. Posteriormente, se muestran las implementaciones e integraciones realizadas sobre los dispositivos que forman la red. Por último, se presenta de forma detallada las comprobaciones y pruebas realizadas sobre el sistema que permitan asegurar el correcto funcionamiento del mismo.

**Palabras clave**—Internet de las Cosas, Redes Inalámbricas, LoRa, LoRaWAN, The Things Network, Chirpstack, Bluetooth Low Energy, Android, Servidores, Interfaz Web.

## I. INTRODUCCIÓN

EN la actualidad, las comunicaciones inalámbricas son parte de la sociedad del siglo XXI. Esto se debe a que las tecnologías siguen avanzando a pasos agigantados y, por tanto, no son las mismas que se conocían hace pocos años. Estos avances nos han permitido comunicarnos e interactuar entre nosotros y con el entorno.

Aunque las tecnologías avanzan, en España aún existen zonas rurales donde no es posible ofrecer cobertura por parte de los operadores. Además, el caso que incentivó este trabajo surge en zonas de Sierra Morena donde un equipo de Protección Civil no tiene cobertura ni conectividad. Por ello, se plantea la idea de crear un red de interconexión móvil para dar soporte a estos equipos de emergencias.

Este trabajo se centra en el uso de diversas tecnologías de comunicación inalámbrica, es decir, se centrará en el uso de Bluetooth y LoRaWAN para la conectividad de los miembros del equipo de emergencias en zonas sin cobertura. Los principales motivos para la elección de estas tecnologías son el uso de las bandas ISM y las características ofrecidas por cada una de estas como envío de datos, bajo consumo energético y rentabilidad.

## II. FUNDAMENTOS TEÓRICOS

### II-A. Bluetooth Low Energy

Bluetooth es una tecnología inalámbrica de corto alcance está basada originalmente en el estándar IEEE 802.15.1 [1].

Bluetooth Low Energy (BLE) es parte de la versión 4.0 del núcleo de especificaciones Bluetooth [2]. Esta versión fue diseñada para optimizar los costes, con bajo ancho de banda, baja potencia y baja complejidad.

Actualmente en el mercado podemos encontrarnos tres versiones diferentes en los dispositivos Bluetooth [3].

- **Bluetooth clásico (BR/EDR):** implementa el estándar inalámbrico clásico Bluetooth.
- **Monomodo (BLE, Bluetooth Smart):** implementa el estándar BLE y puede comunicarse con otros módulos que empleen BLE.
- **Modo dual (BR/EDR/LE, Bluetooth Smart Ready):** implementa ambos estándares, BR/EDR y BLE y es capaz de conectarse con cualquier dispositivo Bluetooth.

En la figura 1, podemos apreciar como realiza la conectividad entre dispositivos Bluetooth con distintas versiones.

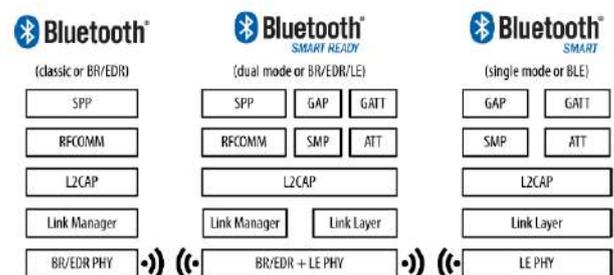


Figura 1. Conectividad entre dispositivos con diferentes versiones [4]

El protocolo Bluetooth Low Energy (BLE) mantiene una estructura de tres partes: controlador, host y aplicación. Estos albergan todas las funcionalidades de su pila de protocolos, como se muestran en la figura 2.

### Controlador

- **Capa Física (PHY):** proporciona la comunicación (modulación y demodulación, transformación A/D, ...).
- **Capa de Enlace (LL):** gestiona las conexiones entre dispositivos BLE mediante la estructura maestro-esclavo.
- **Interfaz de Control de Host (HCI):** se encarga de la comunicación entre el controlador y el host.

### Host

- **Protocolo de Control Lógico y Adaptación de Enlace (L2CAP):** multiplexa los protocolos de las capas superiores y los encapsula en el estándar BLE.
- **Protocolo de Gestión de Seguridad (SMP):** genera las claves de seguridad y el intercambio de las mismas.
- **Perfil Genérico de Atributo (GATT):** define la organización y el intercambio de los datos entre aplicaciones, a través de una arquitectura cliente/servidor.
- **Perfil Genérico de Acceso (GAP):** se encarga del control, detección, conexión y difusión de los dispositivos.

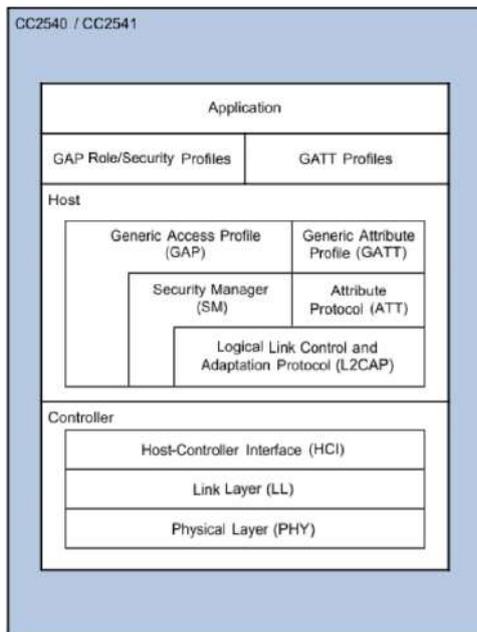


Figura 2. Pila de Protocolos de BLE [5]

### II-B. LoRaWAN

LoRaWAN es una especificación para redes LPWAN propuesta por LoRa Alliance [6] y diseñada para dispositivos de bajo consumo. LoRaWAN define tanto el protocolo de comunicaciones como la arquitectura de red. Así, indica cómo se unen los dispositivos de una red LoRaWAN, gestionando sus canales y parámetros de conexión.

El protocolo LoRaWAN [6] describe las capas: física y enlace lógico; aunque más específicamente, la capa MAC, dentro del sistema de referencia OSI. En la figura 3 se representan las capas mencionadas.

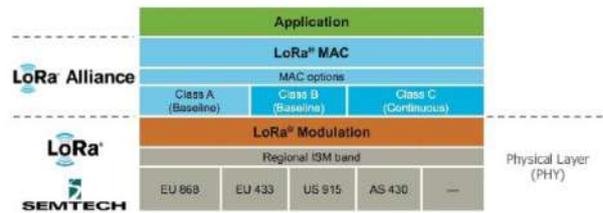


Figura 3. Pila de Protocolos de LoRaWAN [7]

- **Capa Física (PHY):** proporcionan la comunicación y opera en la banda ISM de 433 MHz (Asia), 868 MHz (Europa) o 915 MHz (EEUU).

Esta capa emplea la modulación LoRa para establecer enlaces de comunicación a largo alcance mediante el uso de baja potencia. Además, cabe destacar el parámetro *spreading factor (SF)* que indica el número de canales ortogonales donde se pueden transmitir simultáneamente sin colisiones.

- **Capa MAC:** gestiona el acceso al medio por parte de diferentes dispositivos LoRaWAN, gestionando sus canales y parámetros de conexión. La especificación LoRaWAN define tres tipos de dispositivos según la bidireccionalidad [8]. Los distintas clases de LoRaWAN se encuentran representadas en la figura 4.

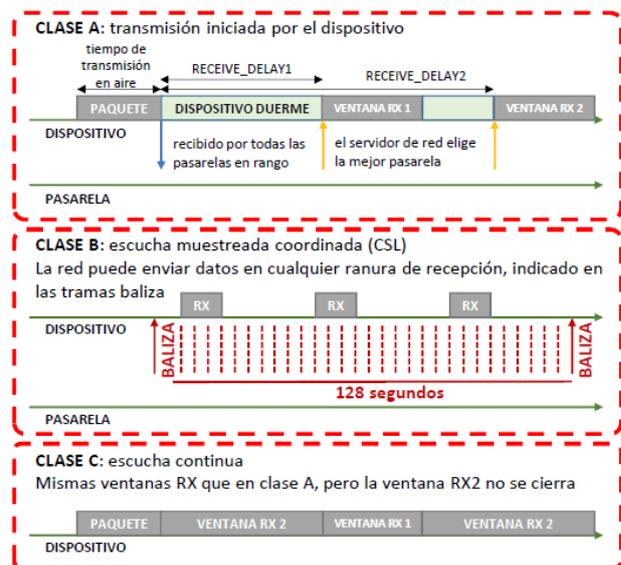


Figura 4. Funcionamiento de las Clases de LoRaWAN [9]

- **Clase A:** los dispositivos únicamente pueden transmitir en ranuras temporales predefinidas.
- **Clase B:** los dispositivos pueden transmitir en ranuras temporales predeterminadas tras recibir una baliza de sincronización o *beacon*.
- **Clase C:** permite la recepción en cualquier momento, excepto cuando se transmite. Esto permite una comunicación con menor latencia, a pesar de un consumo excesivo de la energía.

LoRaWAN define una arquitectura de red y la funcionalidad de seguridad que implementa a varias capas, como se muestra en la parte derecha de la figura 5. Esta red está formada por cuatro tipos de dispositivos [10]:

- **Nodo Final (mota):** dispositivos integrados de comunicación de baja potencia.
- **Gateway:** recibe las transmisiones de las motas y envía datos de respuesta a estas (similar a un bridge).
- **Servidor de Red:** enruta los mensajes entre las motas y los servidores de aplicación y viceversa.
- **Servidor de Aplicación:** se encarga de las solicitudes transmitidas desde los nodos finales.

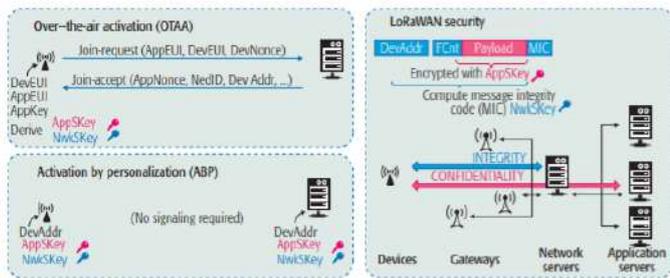


Figura 5. Arquitectura de Red LoRaWAN

Esta red presenta seguridad propia a dos capas, es decir, se emplea seguridad en capa de red y en capa de aplicación:

- **Red:** se encarga de la autenticación de los nodos e incorpora mecanismos de integridad de los mensajes hasta el servidor de red.
- **Aplicación:** incorpora mecanismos de confidencialidad en los mensajes hasta los servidores de aplicación.

Además, existen dos tipos de activación [11]: ABP y OTAA. Estos tipos de activación se muestran en la parte izquierda de la figura 5.

### III. DISEÑO E IMPLEMENTACIÓN

La idea principal del proyecto es desarrollar una red privada con infraestructura propia y móvil, capaz de transmitir información (mensajes y localización) desde cualquier dispositivo móvil al nodo central, incluso si estos nodos se encuentran en zonas fuera de cobertura.

#### III-A. Diseño de la Red

La red diseñada se fundamenta principalmente en la arquitectura de red ofrecida por Chirpstack. Esta arquitectura se implementa como el núcleo de nuestro diseño e incorporamos elementos extras al sistema como la aplicación móvil, un servidor propio, etcétera

Durante el proceso de diseño de la red se analizaron las funcionalidades de cada uno de los dispositivos que conforman la red y las conexiones que se establecen entre cada par de nodos. El diseño final de la red podemos verlo en la figura 6.

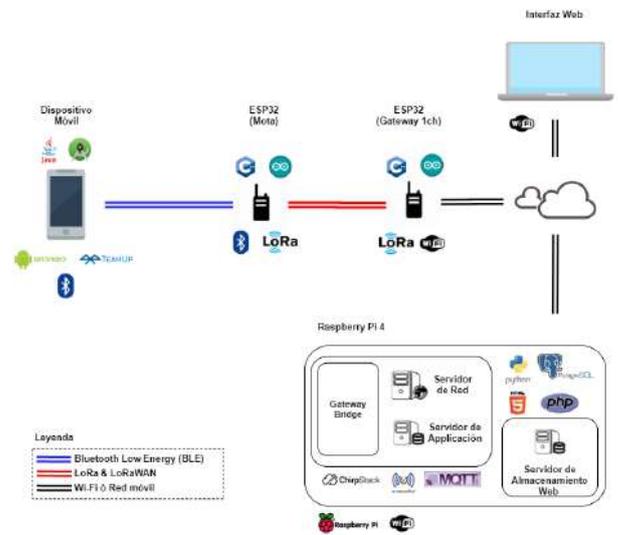


Figura 6. Diseño de la Red de Comunicaciones TeamUp

- **Dispositivo móvil:** el dispositivo móvil se emplea por un miembro del equipo de emergencias. Este dispositivo incorpora la aplicación desarrollada para este proyecto (*TeamUp*).
- **Mota:** es un dispositivo móvil que llevará consigo cada miembro del equipo de emergencias y se encarga de la recepción de mensajes desde el dispositivo móvil mediante el protocolo de comunicaciones BLE y cambio de tecnología del mensaje a LoRa para la retransmisión de los mensajes hasta el gateway o viceversa.
- **Gateway:** se encuentra cerca de nuestro nodo central y se encarga de la recepción de los mensajes LoRa provenientes de las motas y lo retransmite vía Wi-Fi al nodo central.
- **Nodo central:** el nodo central puede permanecer de forma estática o en movimiento dependiendo las circunstancias, siempre que ofrezca la cobertura necesaria al equipo de emergencias. Este nodo central se encarga de mantener actualizados los servidores (chirpstack y web).

Estos servidores se emplean para el almacenamiento de los mensajes que se enviaron desde la aplicación móvil. Además, permite la recepción de solicitudes web. La respuesta a estas solicitudes corresponden con la visualización de una web donde se tendrán distintas funcionalidades como registro de usuarios, visualización de mensajes, etcétera.

- **Interfaz Web:** la interfaz web es una GUI creada para facilitar la visualización de los datos a través de los navegadores web de los dispositivos inteligentes.

Empleando la red completa se permite el envío de los mensajes desde cualquier miembro de equipo de emergencias y la visualización de estos mensajes en formato web. Los dispositivos intermedios retransmiten la información hasta llegar a los servidores que permiten el almacenamiento de los mensajes y su posterior visualización.

### III-B. Diseño e Implementación de la App TeamUp

La aplicación *TeamUp* ha sido desarrollada explícitamente para el proyecto y consta de dos partes: diseño e implementación.

El diseño de la aplicación consiste principalmente en la visualización de las distintas pantallas que la incorporan y en cómo el usuario puede navegar entre ellas. En la figura 7 se muestran estas pantallas y su forma de navegar de forma visual.

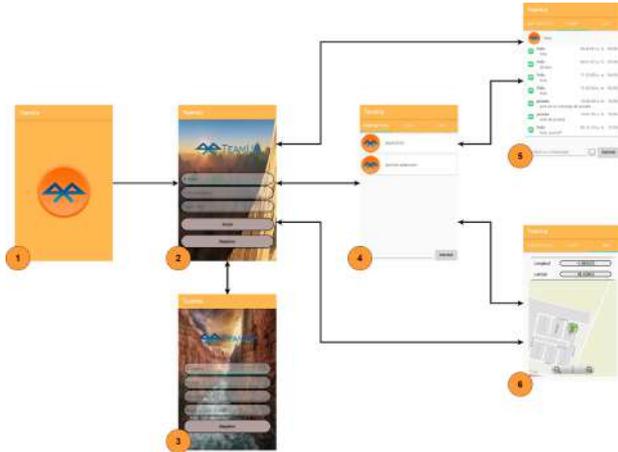


Figura 7. Navegación de la App TeamUp

1. **Portada:** muestra el logo mientras se arranca la aplicación.
2. **Inicio de Sesión:** permite el inicio de sesión de un usuario que esté registrado en la app y, por tanto, almacenado en la base de datos de la red.
3. **Registro:** permite el registro de nuevos usuarios en la aplicación.
4. **Dispositivos:** muestra los dispositivos escaneados que tengan conectividad BLE y permite el envío de mensajes a nuestras motas.
5. **Chat:** muestra los mensajes enviados y recibidos por los distintos miembros del equipo en forma de lista.
6. **GPS:** muestra las coordenadas y un mapa con la ubicación actual del miembro del equipo de emergencias.

La implementación de la aplicación consiste en el desarrollo de las funcionalidades que debe de realizar para pasar entre distintos estados dentro de la misma aplicación. Dicho desarrollo se ha realizado de forma que la aplicación sea compatible para todos los dispositivos que incorporen una versión Android igual o superior a la 4.0.3 (Ice Cream Sandwich).

Por motivos de simplificación, la implementación se subdivide en cinco paquetes:

- **Inicio de Sesión y Registro:** permite el registro de nuevos usuarios y su inicio de sesión. Este paquete implementa las funcionalidades de las tres primeras pantallas del diseño.

- **Principal y Utilidades:** se encarga del control y la gestión de la aplicación.
- **Bluetooth Low Energy:** implementa la conectividad con dispositivos BLE y la gestión de las comunicaciones que emplean este protocolo. Este paquete es usado principalmente en la pantalla 4 del diseño.
- **Chat:** genera chat grupales dependiendo de una clave de grupo. Se encarga del almacenamiento de los mensajes y su visualización. Esta visualización enlaza con la pantalla 5 del diseño.
- **GPS:** se encarga de la localización del usuario y mostrarlo en un mapa (OSM). Este paquete corresponde con la sexta pantalla del diseño.

### III-C. Implementación y Configuración de las Motas

Las motas son elementos que realizan la función de bridge, es decir, hacen el cambio de tecnología entre BLE y LoRa. Estos dispositivos se han implementado empleando dos tipos de empotrados: ESP32-LoRa y Pycom LoPy 4.0.

La primera implementación nos permite reducir los costes económicos y energéticos, mientras que el alcance y capacidad de estos dispositivos están muy limitados debido a ser un dispositivo de clase A. Por otro lado, la implementación sobre la Pycom es un poco más costosa económicamente y energéticamente, pero es un dispositivo de clase C, es decir, permanece en estado de escucha constantemente, mejorando el tiempo de interacción entre usuario y la latencia entre mensajes.

A la hora de la configuración, se han empleado diversos entornos como el IDE de Arduino y Visual Studio Code. Para la configuración en sí, se han empleado librerías existentes para facilitar la labor de la programación.

### IV. IMPLEMENTACIÓN Y CONFIGURACIÓN DEL GATEWAY

Las pasarelas o gateways son nodos intermedios entre los servidores y los usuarios. El gateway se encarga de la recepción de los mensajes LoRa provenientes de las motas y lo retransmite mediante Wi-Fi al nodo central.

En el desarrollo del proyecto se ha optado por la implementación y configuración del gateway dependiendo de tres variables: coste económico y características LoRa (número de canales y tipo de dispositivo empleado). Debido a estas características, se han realizado tres implementaciones distintas:

- **IMST Gateway Lite:** presenta altas capacidades dado que realiza un barrido en los 8 canales disponibles en LoRa. Esta implementación se realiza sobre una Raspberry Pi 1 donde se incluye el *Gateway Bridge* y se configura como se muestra en la figura 8.
- **ESP32 - Single Channel Gateway:** esta implementación se decidió debido a motivos económicos y por el aislamiento social. La implementación del *gateway* monocanal se realizó sobre un ESP32, mientras que el resto de la arquitectura se incorporaron a un host.

La implementación de esta pasarela ha sido empleando una librería desarrollada por Marteen Westenberg que tiene publicado en su repositorio GitHub <https://github.com/kersing/ESP-1ch-Gateway-v5.0>, que nos permite la configuración a distintos niveles: servidores Chirpstack y NTP, características del gateway, spreading factor, detección de la actividad en el canal, etc. Estas configuraciones las vemos en la figura 9.

- LoPy 4.0 - Nano Gateway:** implementa la misma funcionalidad que el *Single Channel Gateway*. La implementación es similar, excepto por el uso del lenguaje Python. En este caso, se han empleado las librerías ofrecidas por la comunidad *Pycom* [12].

```

1 gateway_conf ": {
2   gateway_ID ": " C96000FFFE6A8EA9 ",
3   servers ": [ { " server_address ": "192.168.100.101" , "
4     serv_port_up ": 1700 , " serv_port_down ": 1700 , "
5     serv_enabled ": true } , { "s$
6     ref_latitude ": 38.03403 ,
7     ref_longitude ": -4.08682 ,
8     ref_altitude ": 212 ,
9     contact_email ": " usuario@correo.ugr.es " ,
10    description ": " LORACW01 "
11 }

```

Figura 8. Configuración IMST Gateway Lite

```

1 #define _TTPORT 1700
2 #define _TTNSERVER "192.168.0.25"
3
4 #define _DESCRIPTION "My ESP32 Gateway 868.1MHz SF7"
5 #define _EMAIL "e.fellidelgado@go.ugr.com"
6 #define _PLATFORM "ESP32"
7 #define _LAT 38.03403
8 #define _LON -4.08682
9 #define _ALT 212
10
11 #define _SPREADING SF7
12 #define _CAD 1
13 #define _STRICT_1CH 1
14
15 #define NTP_TIMESERVER "nl.pool.ntp.org"
16 #define NTP_TIMEZONES 1
17
18 #define _DEVADDR { 0x25, 0x00, 0x00 0x00 }
19 #define _APPSKEY { 0x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0
20   x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0x00 }
21 #define _NWSKEY { 0x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0
22   x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0x00 }
23
24 wpa_supplicant - {
25   { "wpa": "" }
26   { "ssid": "TeamUp" , "teanupkey" }
27   { "ssid": "vodafone8778" , "*****" }
28 };

```

Figura 9. Configuración ESP32 Single Channel Gateway

### V. IMPLEMENTACIÓN Y CONFIGURACIÓN DE LOS SERVIDORES CHIRPSTACK

Los servidores Chirpstack son parte del núcleo de la red y se incorporan en la Raspberry Pi 1 (caso de uso de IMST Gateway) o Raspberry Pi 4 (otros gateways). Los servidores son fundamentales en el funcionamiento de la red *TeamUp* dado que se encargan de la recepción de mensajes, descifrado y almacenamiento de los mensajes.

Los red implementada sobre los servidores Chirpstack mantiene una seguridad basada en confidencialidad e integridad. Los mensajes están cifrados desde las motas a los servidores de aplicación y se mantiene la integridad de estos desde las motas a los servidores de red.

La instalación y configuración de los servidores Chirpstack se explican en los siguientes pasos:

- Instalación de los Servidores Chirpstack:** incluye las dependencias (MQTT broker, Redis y PostgreSQL) y configura su correcto funcionamiento.
- Creación de la Organización y Perfil de Servicio :** genera una organización dentro de la estructura *Chirpstack*. Esta organización incluye distintos perfiles de servicio.
- Creación de la Aplicación Chirpstack:** define los dispositivos finales dentro de la estructura *Chirpstack*.
- Creación del Perfil del Dispositivo:** define las propiedades que pertenecen a la organización, e.g. tipo de activación, versión LoRaWAN, etc.
- Configuración de Dispositivos:** asocia los dispositivos a una aplicación y configura las características de las motas (tipo de activación, claves de sesión e identificador del dispositivo).

### VI. IMPLEMENTACIÓN Y CONFIGURACIÓN DEL SERVIDOR PROPIO

El servidor propio mantiene dos funcionalidades claramente separadas. Primero, se encarga de la actualización de la base de datos privada de la red *Teamup* y, segundo, implementa un servicio web para la visualización y gestión de la red por parte de un administrador.

En la actualización de la base de datos, obtiene los mensajes y credenciales de los usuarios de la red mediante el traspaso de información desde la base de datos *Chirpstack*. Esta implementación se ha realizado usando Python y siguiendo el diagrama de flujo de la figura 10.

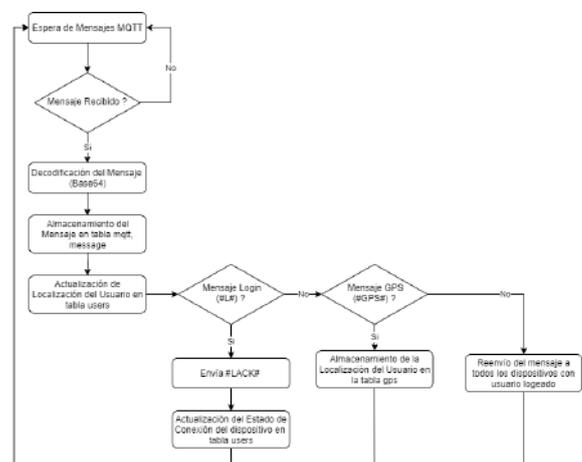


Figura 10. Diagrama de Flujo del Servidor

Por otro lado, el servicio web se implementa de forma que un administrador tenga una serie de funcionalidades. Estas funcionalidades se muestran en la figura 11.



Figura 11. Funcionalidades de la Interfaz Web

### VII. PRUEBAS DE CONCEPTO

Durante el desarrollo del trabajo se realizaron una serie de pruebas para la comprobación del adecuado funcionamiento de la red *Teamup* [13]. Una prueba donde se aprecia el funcionamiento de todo el sistema es *'Inicio de Sesión del Usuario en el Sistema'*. Esta prueba sigue los siguientes pasos:

1. **Puesta en marcha del servidor:** se inicia el servidor Python para actualizar nuestra base de datos privada.
2. **Conexión LoRaWAN:** esperamos a que la mota y el gateway establezcan la conexión LoRaWAN.
3. **Conexión BLE:** se escanean los dispositivos BLE desde la app y se realiza la conexión con la mota.



Figura 12. Inicio de sesión BLE

4. **Envío mensaje Login:** el mensaje de inicio de sesión se transmitirá a través de todos los nodos de la red empleando un formato específico:

#L# [usuario] [contraseña]

5. **Recepción del mensaje Login ACK:** en el momento en el que el mensaje llega al servidor y se procesa, este genera un mensaje de confirmación #LACK.
6. **Mensajes en la Interfaz Web:** el administrador de la red tiene la posibilidad de visualizar los mensajes, entre ellos los de inicio de sesión como aparecen en la figura 13.

UNIVERSIDAD DE GRANADA

Registro de los mensajes recibidos por el servidor de aplicación LoRaWAN:

id	dev_eui	message	time	direction
1	70b3d54994de968f	#L# usuario 1234	2020-06-27 09:33:39.843565+01	uplink
2	70b3d54994de968f	#LACK#	2020-06-27 09:33:40.633869+01	downlink

Figura 13. Mensajes en el Interfaz Web

### VIII. CONCLUSIONES

Este proyecto ha supuesto bastantes dificultades debidas a la complejidad del mismo. Esta complejidad se debe al problema en la interconexión de un equipo de emergencias en zonas sin cobertura.

Principalmente, se realizó un estudio del tecnologías y se seleccionaron las tecnologías inalámbricas BLE y LoRaWAN. Posteriormente, se realizó un diseño de la red e implementó. Dicha implementación se fundamentó en un núcleo de red LoRaWAN (*Chirpstack*) y se desarrollaron desde cero tanto la aplicación móvil el servidor privado y la web.

Finalmente, se realizaron un conjunto de pruebas para verificar del correcto funcionamiento de la red. Se comprobó que la red diseñada es perfecta para zonas sin cobertura dado que emplea tecnologías para ofrecer un amplio rango de cobertura y, además, se comprobó que los usuarios podían enviar y recibir mensajes a través de esta red. Por tanto, podemos afirmar que la red es 100% operativa.

### IX. TRABAJOS FUTUROS

A pesar de tener en pleno funcionamiento la red diseñada, cabe comentar que existen posibles mejoras o ideas a desarrollar aún. Estas se reflejan a continuación:

- Mejora del diseño de la interfaz web.
- Implementación y comprobación de la ruta de un usuario desde la interfaz web, empleando *Google Maps*.
- Diseño de una red mallada (MANET) mediante Wi-Fi como vía complementaria a la red que se encuentra en funcionamiento, de forma que los miembros del equipo de emergencias empleen esta estructura de red en caso de encontrarse cerca unos de otros.
- Análisis del rendimiento, capacidad y coste energético asociado a las tecnologías empleadas en la red actual.

### REFERENCIAS

- [1] IEEE. 'The world's largest technical professional organization dedicated to advancing technology for the benefit of humanity'. Disponible: <https://www.ieee.org/>
- [2] Bluetooth Technology. 'Core Specifications'. Disponible: <https://www.bluetooth.com/specifications/bluetooth-core-specification/>
- [3] Bluetooth Technology. 'Radio Versions'. Disponible: <https://www.bluetooth.com/learn-about-bluetooth/bluetooth-technology/radio-versions/>
- [4] Kevin Townsend Carles Cufí Akiba & Robert Davidson. 'Getting Started with Bluetooth Low Energy: Tools and Techniques for Low-Power Networking', May 2014.
- [5] Texas Instruments. 'CC2540 and CC2541 Bluetooth low energy Software Developer's'. Disponible: <http://www.ti.com/lit/ug/swru271g/swru271g.pdf>
- [6] Documentation of LoRaWAN. Disponible: <https://loralliance.org/about-lorawan>
- [7] Pila de Potocolos LoRa. Disponible: <https://www.aprendiendoarduino.com/tag/lorawan/>
- [8] The Things Network. 'Documentation of Classes of Devices'. Disponible: <https://www.thethingsnetwork.org/docs/lorawan/classes.html>
- [9] Jorge Navarro Ortiz. 'Apuntes Redes Inalámbricas y Movilidad'
- [10] The Things Network. 'Documentation of Architecture of LoRaWAN'. Disponible: <https://www.thethingsnetwork.org/docs/lorawan/architecture.html>
- [11] The Things Network. 'Documentation of Activation'. Disponible: <https://www.thethingsnetwork.org/docs/lorawan/addressing.html>
- [12] Pycom. 'API Reference for Nano Gateway'. Disponible: <https://docs.pycom.io/tutorials/lora/lorawan-nano-gateway/>
- [13] Delgado Ferro, Félix. 'Interconexión de Miembros de un Equipo de Emergencias en Entornos sin Cobertura usando Dispositivos Móviles'. Disponible: [https://wpd.ugr.es/jorgenavarrro/thesis/2020\\_TFG\\_FelixDelgadoFerro.pdf](https://wpd.ugr.es/jorgenavarrro/thesis/2020_TFG_FelixDelgadoFerro.pdf)
- [14] Navarro-Ortiz, Jorge and Sendra, Sandra and Ameigeiras, Pablo and Lopez-Soler, Juan M. (2018,02). 'Integration of LoRaWAN and 4G/5G for the Industrial Internet of Things'. *IEEE Communications Magazine*. volumen 56, págs. 60-67. Disponible: <https://ieeexplore.ieee.org/document/8291115>

# Modeling and optimization for the Cell Switch Off Problem in Ultradense 5G Networks

Autor: Pablo Helio Zapata Cano, e-mail: pablo.zapatacano@gmail.com

Tutor: Juan Francisco Valenzuela Valdés, e-mail: juanvalenzuela@ugr.es

Tutor 2: Pablo Padilla de la Torre, e-mail: pablopadilla@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Abstract**—Ultra-dense networks (UDNs) are recognized as one of the key enabling technologies of the fifth generation (5G) networks, as they allow for an efficient spatial reuse of the spectrum, which is required to meet the traffic demands foreseen for the next coming years. However, the power consumption of UDNs, with potentially hundreds of small base stations (SBSs) within each macrocell, is a major concern for the cellular operators, and has to be properly addressed prior to the actual deployment of these 5G networks. A widely accepted strategy for addressing this issue lies in the selective deactivation of SBSs, but without compromising the QoS provided to the User Equipments (UEs), known as the Cell Switch-Off (CSO) problem. In this work, the CSO problem is formulated as an optimization problem and addressed with various multiobjective genetic algorithms (GAs) in different manners. Furthermore, specific operators are proposed in order to introduce knowledge about the problem into the algorithms, making them more efficient.

**Key words**—5G, Cell switch-off, Multi-objective Optimization, UDN, metaheuristics, parallelism

## I. INTRODUCTION & STATE OF THE ART

The latest mobility reports published by Ericsson [1] and Cisco [2] are just confirming the predicted previsions about the increasing global mobile data traffic. The former, updated up to the first quarter (Q1) 2019, announces that such a traffic has grown by 82% between Q1 2018 and Q1 2019, reaching more than 28 exabytes per month and more than 6.0 billion mobile broadband subscriptions globally. On its part, Cisco has forecasted this amount of traffic to increase to 77 exabytes by 2022. Under this scenario, both public and private initiatives started to develop the fifth generation (5G) of cellular systems. Three paradigms has been clearly identified to reach the challenging design requirements and expected performance indicators of 5G networks [3]: moving to millimeter wave (mmWave) spectrum to use larger bandwidths, enhancing spectral efficiency via multi-antenna transmission (massive, collaborative MIMO), and finally increasing spatial reuse through network densification [4].

In order to reduce the propagation loss of mmWave communications, improve the signal to interference plus noise ratio and reduce the latency, 5G networks require the deployment of a large number of small base stations (SBSs), which are close to the mobile users [4]. They are named ultra-dense networks (UDNs) [5], [6]. However, these dense deployments come with a considerable increase in the power consumption of the system as SBSs are the most consuming device of the network (up to 80%), regardless of its load.

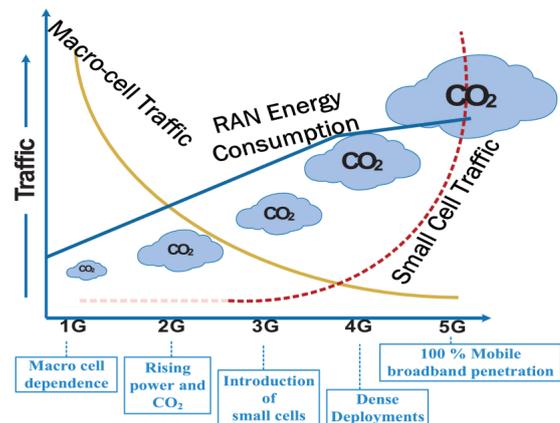


Fig. 1: Trends in traffic, energy consumption and CO<sub>2</sub> emissions, for different generations of wireless communication.

This major issue could be addressed at different network levels [7], but one of the most promising one, to the point of being standardized by the 3rd Generation Partnership Project association [8], lies in switching off a number of SBSs when they are underutilized during low traffic demand periods. This is known as the Cell Switch-Off (CSO) problem [9]. CSO problems can be classified into two main categories [10] online and offline CSO. Whereas online approaches are performed in real time (or in a few minutes basis), offline ones determines the set of active cells for longer time scales (from dozens of minutes to hours) in which the network conditions are assumed to be constant. Our work deals with offline CSO problems unless otherwise specified. However, the decision problem that determines which subset of SBSs should be switched off in offline CSO is a NP-complete problem [11]. It also has to account not only for the energy consumption of the UDN, but also for any indicator of the QoS provided to the users. The CSO problem has also been addressed from different perspectives, ranging from clustering techniques or game theory, to its formulation as an optimization problem, tackling it with both simple heuristics and metaheuristics.

## II. SYSTEM MODEL

This section aims to describe the model on which all improvements and experimentations are based. A brief description of the two main algorithms used in the experimentations is also provided.

### A. The ultradense network

The target service area has a dimension of 500 square meters (implemented through a discretized 100x100 points grid). In this space, the power of the signal is assumed to be constant. With the aim of emulating different propagation conditions, ten different regions have been set, characterized by an attraction point. On this basis, the propagation parameters of the region are generated randomly. These are, among others, the path loss and the channel type of the region. Once the ten propagation regions have been initialized, a Voronoi tessellation based on them is computed. In this way, the closest region is found for each point of the grid. Four different type of cells of different sizes are considered: femtocells, picocells, microcells and macrocells (see Fig 2). In this base model, the criteria for users to connect to a certain BS is to choose the one with the highest SINR, regardless of its type.

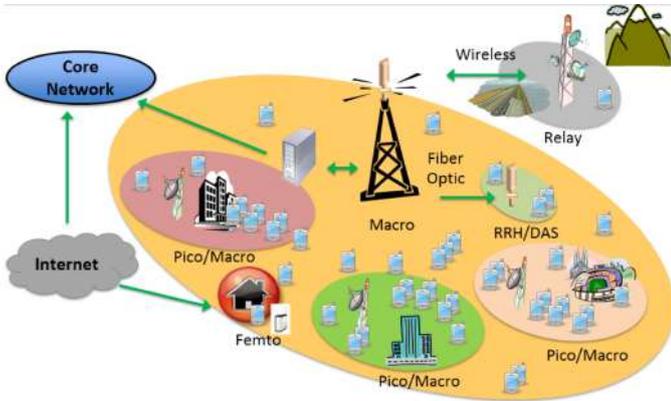


Fig. 2: HetNet scheme used in the model

Moreover, two subtypes of micro, pico and femtocells are also defined, summing up 7 cell types. All serving BSs of all types have a transmitting power  $P_{tx}$  of 750mW. The operation frequency depends also from the celltype, being the smaller cells (femto1 and femto2) the ones operating at the highest frequencies, and the macro cell the one operating at the lower one.

Within this work, a number of new and improved calculations have been added to the initial model, making it more real and accurate. These are:

- **New power consumption model:** A new aggregated power consumption model has been implemented, considering the power consumption of the SFP device and the intra- and interbackhaul powers.
- **Provisioning stats:** Three metrics have been considered in order to track the over- and underprovisioning of resources at the transmitter.
- **3D and dynamism:** The model is now three-dimensional and the mobility of the users is considered.
- **Beamforming and MIMO:** The power of the antenna varies now with the angle of transmission. Moreover, multiple antennas at the transmitter and the receiver are considered in a MIMO model.

### B. Algorithms

In the following, a classification and a description of the algorithms used in this work is provided:

1) *NSGAI*: The Non-Dominated Sorting Genetic Algorithm II, NSGA-II, was proposed by Deb *et al.* [12]. It is a genetic algorithm based on generating a new population from the original one by applying the typical genetic operators (selection, crossover, and mutation); then, the individuals in the new and old population are sorted according to their rank, and the best solutions are chosen to create a new population.

2) *MOCe*ll: The Multi-Objective Cellular Genetic Algorithm, MOCe, is a cellular genetic algorithm (cGA) [13] that includes an external archive to store the non-dominated solutions found so far. The archive is bounded and uses the crowding distance of NSGA-II to keep diversity in the Pareto Front. The selection is based on taking an individual from the neighborhood of the current solution and another one randomly chosen from the archive. After applying the crossover and mutation operators, the new offspring is compared to the current one, replacing it if better; if the solutions are nondominated, the worst individual in the neighborhood is replaced by the current one. In these two cases, the new individual is added to the archive.

## III. EXPERIMENTATION

This chapter elaborates on the experimentation part of the work. With the aim of structuring this part, three different experimentations have been carried out. In each of them, particular features of the model are tested. Every of the parts also has a specific goal, i.e., the same problem is faced with a different configuration in each case, which also leads to different contributions.

### A. Hybrid metaheuristics applied to the Cell Switch Off problem in 5G deployments

In this experiment, the so called CSO operator is tested for the first time. The results obtained in this simulation have been published in [14].

1) *Motivation and contributions. The CSO problem.:* Within this experiment, a specific operator for the CSO problem is included in both the NSGA-II and MOCe, algorithms. Despite being apparently very simple, the operator turns to show an considerable search intensification capacity, which leads to an reasonable application of it. For instance, if all cells with no connected users are switch off, the algorithm will not be able to evolve solutions that reallocate users from a BS to an empty one, since all empty BSs would have been switched off when applying the operator. A pseudocode of the operator is shown below:

---

**Algorithm 1** Pseudocode of the "no connected users" operator

---

```

1:  $C \leftarrow \text{cells}(UDN)$ 
2: for  $c$  in  $C$  do
3:   if  $(\text{ConnectedUEs}(c) == 0)$  then
4:      $\text{switch-off}(c)$ 
5:   end if
6: end for
    
```

---

This allows us switching off cells that are not assigned to any user. Its application, as well as the crossover and mutation, depends on an application rate,  $CSO_r$ , that has been set to different values in the experimentation.

2) *Results:* In this section the experimentation results are shown, using the hypervolume and the attainment surfaces as indicators. The goal is to measure the combination of the NSGA-II and MOCell algorithms and the CSO operator. In order to do that, the operator has been applied using two different application rates: 0.01 and 0.1.

The first conclusion that one can extract from the results is that the hybridization with the CSO operator improves the search of the algorithms. None of the canonical algorithms shows better results than the hybrid version in none of the scenarios. This is a really promising result, since this kind of hybrid approximation is not common in this field and rarely appear in the literature. Secondly, it can be seen that the hybrid version with a lower application rate (0.01 vs 0.1) reach better approximations to the Pareto front. This is because of the effect that the operator has on the diversity of the population.

Analyzing the results of the second indicator (attainment surfaces), it can be observed that, the higher the UE density, the higher the total capacity of the network. This was expected as a consequence of densification. As the density of the scenario increases, the mean approximations to the reference front of the hybrid versions are increasingly separated from those of the canonical versions. The fact is that hybrid versions sample more frequently zones with few active BSs (left part of the representation in the Fig 3). However, canonical versions obtain non-dominated solutions with a high number of active stations, which leads to an increase of the power consumption. Despite consuming more, the solution reached by the canonical algorithms imply a higher value of the capacity i.e. right-top part of the representation in Figure 3.

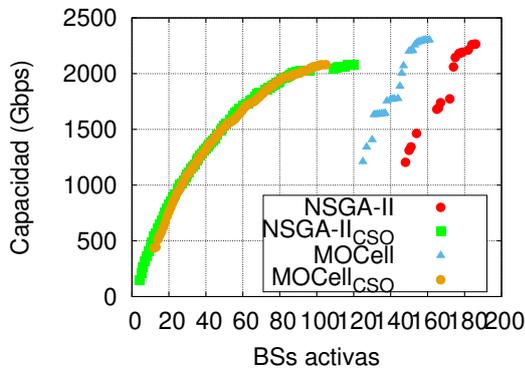


Fig. 3: Results of the attainment surfaces for a dense configuration of the network.

This is provoked because the CSO operator is having impact only on one of the two objectives (the consume), neglecting the other (capacity).

3) *Conclusions:* As general conclusion, a multi-objective optimization problem has been formulated, minimizing the consume (number of active BS) and maximizing the capacity of the network (GBps in the downlink) at the same time. Furthermore, an specific local search operator, called CSO, has been proposed and integrated in both the NSGA-II and MOCell algorithms. The hybrid versions (the ones that include the CSO operator) reach better approximations to the Pareto Front. The results show a considerable improvement in the HV indicator. Regarding the attainment surfaces, they showcase

that the operator has centered the search in zone with a low number of active BSs and, consequently, with a low capacity (areas in the left). However, the canonical versions obtain non-dominated solutions with a large number of operative BSs (which implies a higher consume), but also with a higher capacity (areas in the top-right). This is caused because the CSO operator just improves one of the objectives: the consume, obtaining worse results in terms of capacity.

### B. A robust approach to the cell switch-off problem in 5G ultradense networks

Through this experiment, a robust version of the CSO problem is proposed. Note that the obtained results have been published in [15].

1) *Motivation and contributions:* A common assumption in CSO implementations is that the traffic demand in the network is usually estimated [16]. As so, the optimization algorithms are aimed to look for solutions based on such estimates, which might lead to an overfitting when having inaccuracies, causing a poor performance of the network. In this approach the aim is to make the solutions to the CSO problem robust [17], i.e., to add them some degree of insensitivity to disturbances in the estimation of the traffic demands. To do so, additional objectives that measure such robustness are introduced, resulting in the definition of the Robust CSO (RCSO) problem.

In order to take advantage of a higher number of computing elements (processors, cores, etc.), a parallelization of the NSGA-II algorithm, named as mwNSGA-II, has been implemented, following the classical master-worker paradigm. This new version breaks down the synchronization requirements imposed by the evolutionary loop of the sequential algorithms by generating as many solutions as workers available, and incorporating the evaluated solutions into the population regardless of the generation it was created.

2) *The Robust CSO Problem:* Since measuring the robustness of a solution is a mathematically and challenging task, an approximated procedure based on a probabilistic sampling has been used: Monte Carlo integration. For each solution, we have  $H$  different objective values from which its robustness has to be computed. This is undertaken by extending the two-objective CSO problem to a four-objective RCSO problem that aims at minimizing the average power consumption and maximizing the average capacity over the  $H$  samples, but also minimizing their two variances, as indicators of robustness.

3) *Results:* This section has been further structured into three different subsections for a better organization of the discussion of the results:

- **Parallel performance:** In order to show the performance reached by the parallel execution of mwNSGA-II, the wall-clock time, the accumulated run time of all the workers, the parallel efficiency, and the number of workers involved in the computation have been computed over 5 independent runs. The mwNSGA-II algorithm has been able to reduce the computational time of the search from dozens of days to roughly 1-2 hours. The values of the parallel efficiency attained by mwNSGA-II indicate how the algorithm takes full advantage of the computational power provided by the platforms, reaching a parallel efficiency over 95%.

- Type of solutions of the RCSO problem:** The four objectives of the RCSO problem are represented by boxes (see Figure 4), in which the position of the center of the boxes is defined by the average values of the power,  $\mu_{f_{Power}}$ , and the capacity,  $\mu_{f_{Cap}}$ , whereas the width and the height are set up with the standard deviation,  $\sigma_{f_{Power}}$  and  $\sigma_{f_{Cap}}$ . The idea is that, the higher the area, the lower the robustness.

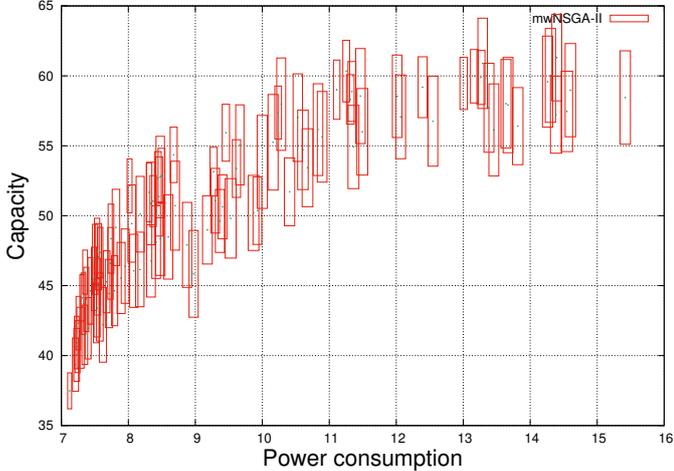


Fig. 4: A Pareto front of the HH instance of the RCSO problem.

It can be seen that solutions with a low power consumption (the left hand side of the plot) tend to be more robust (the area of the squares are smaller) and, as long as the UDN has to provide higher capacity to the UEs (right part), solutions become more sensitive to the traffic data (less robust). That is, the UDN configurations that leave more SBS switched on are more sensitive to changes in the traffic demands. A second interesting finding is that solutions are much more robust with respect to the power consumption than with respect to the capacity. This is reflected as the width of the boxes is smaller than their height. In any case, this set of solutions provides the network operator with a highly valuable information on how rather close configurations may better react to unexpected traffic demands.

- Sensitivity analysis:** In order to evaluate the impact of introducing robustness in the CSO problem formulation, that is, the definition of the RCSO problem, this section develops a sensibility analysis of the solutions obtained by a sequential run of NSGA-II that only considers one traffic demand scenario. To do so, a completely different traffic demand pattern has been generated. In this way, users follow the well known random waypoint mobility model. Under this new demand, the robust solutions, i.e., the switching off plan of small base stations, of the RCSO problem computed by mwNSGA-II and those in which robustness is not taken into consideration.

	CSO	RCSO
Average power	1.55e-02	9.67e-03
Average capacity	1.39e+01	2.62e+01

Table I: Average power consumption and capacity over a randomly generated mobile traffic pattern in the HH instance

Table I includes the average power and average capacity

of 100 time steps of the UEs moving within the service area of the HH instance. Here, the conclusions are also quite direct. The solutions provided by the RCSO problem reach not only a lower power consumption, but also a higher capacity. That is, if the problem of switching off SBSs of UDNs is addressed from a robust optimization approach, unexpected or miss-estimated traffic demands can be better serviced by the network, thus increasing the QoS and also the user experience.

4) *Conclusions:* This work encompasses the task of how the power consumption and the capacity of UDN networks vary when the traffic demand estimates have disturbances. In order to provide the solutions with robustness, the original bi-objective CSO problem has been extended to the four-objective RCSO problem, which provoked a considerable increase in the computation time, because each solution has to be evaluated over  $H$  different traffic demands. To address this issue, this work has used a parallel version of the NSGA-II algorithm, mwNSGA-II, that follows a master-slave approach. The results over two different instances and on two different parallel computing platforms have provided very promising solutions. Indeed, the algorithm is able to reach more than 95% of parallel efficiency, and is able to reduce the total expected computation time of the algorithm from 42.3 days to 2.1 hours. As to the analysis of the solution quality, the solutions reached are more robust in regions of low power consumption configurations of the search space. The larger the capacity of the network, the lower the expected robustness. A sensibility analysis has also shown that the solutions of the RCSO problem are less sensitive to disturbances than those computed by the original CSO formulation.

### C. Approaching the cell switch-off problem in 5G ultradense networks with dynamic multi-objective optimization

In this experimentation the CSO problem is again addressed, but this time approaching the problem from a dynamic perspective. A paper including the work done within experimentation has been accepted by the Future Generation Computer System Journal (Q1) [18].

1) *Motivation and contributions:* In order to face the QoS optimization objective within the offline CSO framework, most of the existing works use a service demand that is considered also to be fairly constant during a given time span [11], without passing information from one optimization stage to the next one. This is precisely one of the contributions of the work: approaching the CSO problem from a dynamic multi-objective optimization perspective, in which the CSO problem for a traffic pattern at period  $t$  is optimized, but starting from the search experience of the algorithm in the previous period  $t - 1$ , and these traffic patterns change from one period to another. The new formulation of the problem has been named as the dynamic CSO (DCSO) problem. A new restart operator is proposed for the DCSO problem, showing very promising results over a wider variety of scenarios. Both NSGA-II and MOCeII algorithms have been endowed with it, resulting in DNSGA-II-ACR and DMOCeII-ACR algorithms. The behavior of the ACR operator has been fully characterized by evaluating 80 configurations in a systematic experimental setup.

2) *The dynamic CSO (DCSO) problem:* In this dynamic version of the CSO problem, named DCSO, the time comes into play, so the problem objectives defined for the static problems in the section above now depend on the period of time they are computed. This is one of the major contributions of our approach, as the existing related works that face the CSO problem consider that each of these periods is an entirely new instance, which is addressed independently by a MOEA [11]. We introduce here the concept of *epoch* within the CSO framework. An epoch is one of such a time interval of a predefined duration, which may range from dozens of minutes to few hours, in which the positions of UEs (i.e., the traffic demands) is assumed to be constant. After the time of an epoch is elapsed, UEs move to a new position in the service area using the well known Random Waypoint Mobility model.

In summary, the DCSO problem has to minimize the power consumption of the UDN network, maximize the capacity provided to the UEs, and minimizing the number of handovers induced by the new switching on plan for a new epoch.

3) *The ACR operator:* The rationale in the design of this operator is to exploit the locality and continuity of the movements of users within the service area of the UDN. The action of the ACR operator is to switch on the adjacent or neighboring SBSs of  $j$  with one main goal: enabling the algorithms to better track the approximated Pareto front by gradually activating SBSs in the areas where users may potentially move to. This introduces highly valuable genetic material in the solutions that will be passed on to the next epoch. Additionally, this strategy also provides the users with alternative SBSs to be associated with, thus balancing the load and increasing the capacity objective.

---

**Algorithm 2** Pseudocode of the ACR operator
 

---

```

Input 1:  $P$  //A population of solutions
Input 2:  $r_p$  //The application rate on  $P$ 
Input 3:  $r_{on}$  //The application rate on each SBS
Input 4:  $l_{on}$  //The length of the activation ratio

1: for  $\forall p \in P$  do
2:   Select  $p$  with probability  $r_p$ 
3:   for  $\forall b \in \text{SBSwithUsersAssociated}(p)$  do
4:     for  $\forall n \in \text{Neighborhs}(p, l_{on})$  do
5:       Select  $n$  with probability  $r_{on}$ 
6:       SwitchOn( $n$ )
7:     end for
8:   end for
9: end for
  
```

---

The pseudocode of the ACR operator is included in Algorithm 2. It is a randomized procedure that works on a population of solutions  $P$ , and does have three configuration parameters:  $r_p$  indicates the percentage of solutions from  $P$  that undergo to SBS activation,  $r_{on}$  determines the percentage of adjacent SBSs of a given SBS  $p$  that will be activated and, finally,  $l_{on}$  is the range of action of the operator for each SBS  $p$ , measured in terms of grid points of the discretized service area. The operator chooses a subset of solutions from  $P$  and then, for all the SBSs of each of these solutions having users associated (line 3), which could potentially move in the next epoch, it selects the adjacent SBSs within a range defined

by  $l_{on}$  (function  $\text{Neighborhs}(p, l_{on})$  on line 4) and switches on those that are deactivated. The possibly modified set of solutions  $P$  is the output of the method.

4) *Results:* The first conclusion to be highlighted is that any ACR-enabled dynamic algorithm outperform the NRC versions. This means that the genetic material that the ACR operator passes on to the next epoch is very helpful to approximate the newly induced Pareto front defined by the new users positions. If we analyze the results by the ACR parameter, it can be seen that when larger portions of the final set of non-dominated solutions at epoch  $e$  are passed on to epoch  $e + 1$ , the two algorithms are able to better react to the change in the traffic demand, and thus reaching better approximations to the Pareto fronts.

The value of parameter  $r_{on}$  is much less important on the behavior of the ARC versions of the algorithm. It is also remarkable that differences in the HV values are very tight in all the aggregated executions. This could be explained by the high density of deployed SBSs in all the considered instances, and the overlapping among the different neighborhoods, which ends up propagating a generalized activation of SBSs in localized regions of the networks where a traffic demand (users) exists.

Finally, the impact of  $l_{on}$  follows a different trend, as the aggregated HV values get decreased when the value of this parameter gets bigger. With higher values of this parameter, the operator activates very large numbers of SBSs in the UDN, driving the algorithms in the next epoch to restart the search from an network operating at full power. In the end, larger values of  $l_{on}$  do not allow to properly track the mobility of the users for the next epoch.

5) *Comparison with other dynamic MOEAs:* The final analysis of the results now targets actual algorithms, not averaging over a set of different configurations. Two different configurations of DNSGA-II-ACR and DMOCcell-ACR has been included in the comparison: the configuration that perform the best considering the 9 UDN scenarios in Fig 5.

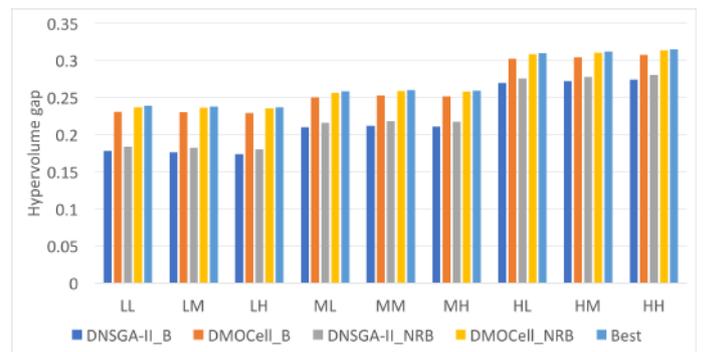


Fig. 5: Gap of the HV value for the different dynamic MOEAs with respect to the NRC version.

The main conclusion that can be drawn from these tables is that the ACR versions of the algorithms have always better approximated the Pareto fronts of the 9 considered instances (higher HV values). Differences are significant with respect to the A version, and a bit tighter in the B versions, in which the configuration with  $r_p = 0.95$  has also performed quite well. In order to better show the benefits of the ACR

operator, Figure 5 displays the gap between the HV value of the NRC version, which serves as a baseline, and the different dynamic MOEAs that incorporates a reaction-to-change mechanism. It can clearly be seen that the two last columns in each group, which correspond to the ACR-enable versions of the algorithms, have the larger differences. Given the normalization procedure followed to compute the HV values, these minimal differences induce substantial changes in the approximated fronts.

6) *Conclusions:* This work has approached the Cell-Switch Off problem in ultra-dense 5G networks from a dynamic evolutionary multi-objective optimization perspective, endowing the algorithms with mechanisms to react to the change in these patterns so as to better track the new Pareto front of the next epoch. In order to improve such tracking, a new restart operator based on exploiting the continuous mobility of the users within a cellular network has been devised. This operator, called Adjacent Cell Restart (ACR), can be included in any dynamic MOEA. The results over nine different UDN scenarios has shown that ACR clearly aids the search of these two algorithms to better approximate the new Pareto fronts induced in each epoch, with respect to both algorithms not reacting to change and existing dynamic MOEAs without the ACR operator.

#### IV. GENERAL CONCLUSIONS AND FUTURE WORK

Within this work, a considerable set of improvements and new features have been added to the modeling of an ultradense 5G network, starting from a base model. In addition, the performance of this network has been tested and simulated using genetic algorithms combined with specific operators. As a major conclusion, a new and more realistic model has been created, being able to emulate real future 5G scenarios. Furthermore, a powerful optimization tool has been set out, adding knowledge of the problem through the introduction of CSO specific operators. Moreover, new versions of the optimization have been designed, adding robustness and dynamism to the problem.

The achievements described above clearly open promising lines to be further explored. On the one hand, the modeling of the network can be further extended by introducing data from real 5G scenarios. On the other hand, more algorithms and operators can be incorporated to the optimization framework. In the long run, also artificial intelligence techniques can be used for optimizing the performance of the network.

#### ACKNOWLEDGMENTS

This work would not have been possible without the constant, professional and close help of the two tutors, Juanfra and Pablo, but also Paco and Ángel. I also acknowledge the Smart Wireless Technologies Group and the Universities of Granada and Málaga for all the help and assistance provided. In the human and moral aspect, I would like to highlight the unconditional support of my family and friends, that have been my side in every moment throughout the realization of this work.

#### REFERENCES

- [1] Ericsson, "Ericsson mobility report," <https://www.ericsson.com/en/mobility-report/reports/q4-update-2018>, 2018, accessed June 8, 2019.
- [2] Cisco, "Global mobile data traffic forecast update, 2017–2022 white paper," <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html>, 2019, accessed June 8, 2019.
- [3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [4] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in Cellular Systems: Understanding Ultra-Dense Small Cell Deployments," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 4, pp. 2078–2101, 2015.
- [5] X. Ge, S. Tu, G. Mao, C. X. Wang, and T. Han, "5G Ultra-Dense Cellular Networks," *IEEE Wireless Communications*, vol. 23, no. 1, pp. 72–79, 2016.
- [6] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-Dense Networks: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.
- [7] P. Gandotra, R. K. Jha, and S. Jain, "Green Communication in Next Generation Cellular Networks: A Survey," *IEEE Access*, vol. 5, pp. 11 727–11 758, 2017.
- [8] 3GPP, "3gpp, small cell enhancements for e-utra and e-utran—physical layer 47 aspects," <http://www.3gpp.org/ftp/Specs/html-info/36872.htm>, 2014, 2014.
- [9] M. Feng, S. Mao, and T. Jiang, "Base Station ON-OFF Switching in 5G Wireless Networks: Approaches and Challenges," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 46–54, 2017.
- [10] T. Beitelmal, S. Szyszkowicz, G. David González, and H. Yanikomeroglu, "Sector and site switch-off regular patterns for energy saving in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2932–2945, 2018.
- [11] D. Gonzalez G., J. Hamalainen, H. Yanikomeroglu, M. Garcia-Lozano, and G. Senarath, "A Novel Multiobjective Cell Switch-Off Framework for Cellular Networks," *IEEE Access*, vol. 4, pp. 7883–7898, 2016.
- [12] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182 – 197, 2002.
- [13] A. J. Nebro, J. J. Durillo, F. Luna, B. Dorronsoro, and E. Alba, "Mocell: A cellular genetic algorithm for multiobjective optimization," *International Journal of Intelligent Systems*, vol. 24, no. 7, pp. 726–746, 2009. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/int.20358>
- [14] P. H. Zapata-Cano, F. Luna, J. F. Valenzuela-Valdes, A. M. Mora, and P. Padilla, "Metaheurísticas híbridas para el problema del apagado de celdas en redes 5G," *Congreso Español de Metaheurísticas, Algoritmos Evolutivos y Bioinspirados*, 2018.
- [15] F. Luna, P. H. Zapata-Cano, J. F. Valenzuela-Valdes, and P. Padilla, "A robust approach to the cell switch-off problem in 5G ultradense networks," *International Conference on High Performance Computing & Simulation*, 2019.
- [16] M. Katsigiannis, *A demand model for mobile data traffic in the 5G Era: Case of Finland*. IGI Global, 2018.
- [17] Z. He, G. Yen, and Z. Yi, "Robust multiobjective optimization via evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 2, pp. 316–330, 2019.
- [18] F. Luna, P. H. Zapata-Cano, J. C. Gonzalez-Macias, and J. F. Valenzuela-Valdes, "Approaching the cell switch-off problem in 5G ultra-dense networks with dynamic multi-objective optimization," *Future Generation Computer Systems*, 2019.



**Pablo Helio Zapata Cano** Pablo Zapata received the BSc degree in Telecommunications Engineering from the university of Granada (Spain). Currently he is a SMARTNET MSc scholar at TelecomSud Paris, where he is recipient of an Erasmus Mundus grant. Since 2018 he has been also active in research as a graduate research assistant at the University of Málaga, publishing several articles in national and international conferences. His research interests include 5G Ultradense Networks, green communications and optimization techniques.

# Network Slicing in 5G Systems: Key Enablers, Architectures, and Performance Evaluation

Autor: José Antonio Ordóñez Lucena, e-mail: jordonez@ugr.es  
Tutor: Pablo Ameigeiras Gutiérrez, e-mail: pameigeiras@ugr.es  
Tutor: Juan José Ramos Muñoz, e-mail: jjramos@ugr.es  
Master's Degree: Telecommunications Engineering  
Department of Signal Theory, Telematics, and Communications  
University of Granada

**Abstract**—The fifth generation of mobile communications is anticipated to open up innovation opportunities for new industries such as vertical markets. However, these verticals originate myriad use cases with diverging functional and performance requirements that future 5G networks have to efficiently support. Network slicing may be a natural solution to simultaneously accommodate over a common network infrastructure the wide range of services that vertical-specific use cases will demand. In this article, we present the network slicing concept, with a particular focus on its application to 5G systems. We start by identifying the requirements in network slicing, and discussing the key capabilities that may enable us to satisfy them, including the integration of Software Defined Networking (SDN) and Network Functions Virtualization (NFV) technologies and their required interplay. Following that, we propose a slicing architecture based on SDN and NFV, and we show that the combined use of both technologies may enable the fulfillment of the slicing requirements. The feasibility of this proposal will be checked using the Open Source MANO (OSM) software stack, and its performance will be evaluated through a slicing-ready experimental setup. Finally, we summarize the main conclusions derived from this work.

**Keywords**—5G Systems, Network Slicing, SDN, NFV, OSM

## I. INTRODUCTION

5G systems are expected to satisfy the consumer, service and business demands of 2020 and beyond. One of the key drivers for 5G systems is the need to support a variety of vertical sectors such as manufacturing, automotive, healthcare, energy, and media&entertainment [1]. Such verticals originate very different use cases, which impose a much wider range of requirements than existing services do nowadays. Today's 4G networks, with their "one-size-fits-all" architectural approach, are unable to address the diverging performance requirements that verticals impose. To accommodate vertical-specific use cases along with increased demands for existing services over the same network infrastructure in a cost-efficient manner, it is accepted that 5G systems will require architectural enhancements with respect to current deployments.

Network softwarization, an emerging trend which seeks to transform the networks using software-based solutions, can be a potential enabler for accomplishing this. Through technologies like Software Defined Networking (SDN) and Network Functions Virtualization (NFV), network softwarization can provide the programmability, flexibility, and modularity that is required to create multiple logical networks, each tailored for a given use case, on top of a common network. These logical

networks are referred to as *network slices*. We define network slices as end-to-end logical networks running on a common underlying (physical or virtual) network, mutually isolated, with independent control and management, and which can be created on demand. Such self-contained networks must be flexible enough to simultaneously accommodate diverse business-driven use cases from multiple players on a common network infrastructure (see Fig. 1).

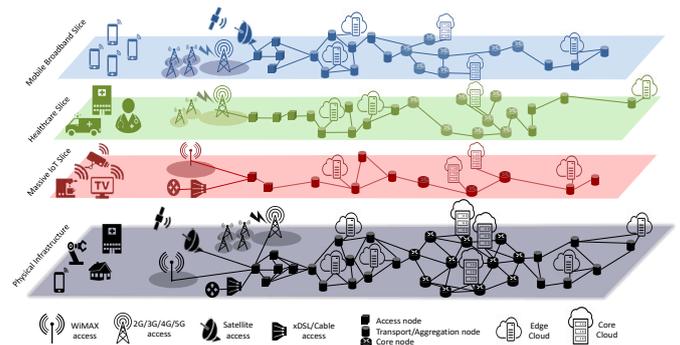


Figure 1. 5G network slices running on a common multi-access, multi-vendor network [2]. Each slice accommodates a particular use case.

In this paper, we provide a comprehensive study of the network slicing concept in 5G. Then, we propose a 5G slicing architecture by integrating both SDN and NFV technologies into a unified management and orchestration framework that fully exploits slicing benefits. Then, we study the feasibility of this solution using the Open Source MANO (OSM) stack, a software environment that brings together open source SDN and NFV tools. Furthermore, we evaluate the performance of our solution using an experimental setup that illustrates the difficulties that slicing imposes in terms of resource orchestration. Resource orchestration consists of optimally dispatching resources across the different slices that concurrently run on a common shared infrastructure, trying to maximizing their performance while minimizing their resource usage.

This work has reached an outstanding international dissemination in the research community. The proposed network slicing architecture is overtly analyzed in a paper [2] that **the student, along with its research colleagues and in collaboration with Telefónica I+D, has published in the IEEE Communications Magazine** (Impact Factor (2017):

10.435; Cat:Telecommunications; Ranking 2/89; Q1). This architecture has also been sent to the European Telecommunications Standardization Institute (ETSI), the main NFV standardization body. **The work has received a very positive welcome from ETSI, that has approved our proposal for standardization in [3].**

The remainder of this paper is organized as follows: Section II provides a background on key slicing enablers. Section III describes our architectural proposal. Section IV validates this proposal using OSM, and Section V evaluates it, with a slicing-ready scenario. Finally, Section VI highlights some conclusions.

## II. TECHNICAL BACKGROUND FOR NETWORK SLICING

### A. Network Slicing Requirements

The network slice definition given in Section I clearly lay the foundation of network slicing requirements:

- **Created on demand:** slices need to be created in short timescales to accommodate use cases with great agility
- **Self-contained:** A slice shall be able to operate autonomously, without human intervention. To be intelligent platform forms complete at their own, each slice shall have its own architecture, policy-based management tools, engineering mechanisms, optimization algorithms, and protocols, all them customized to address the (potentially time-varying) particularities of the use case it accommodates.
- **Isolation:** To run concurrently on top of a shared network, slices need to be mutually isolated in terms of performance, management, and security and privacy. *Performance isolation* ensures that service-specific performance requirements are always met on each slice, regardless of the congestion and performance levels of other slices. *Management isolation* means that management actions have to be performed in a per-slice basis. Finally, *security and privacy isolation* means that each slice has independent mechanisms for protecting integrity and confidentiality, averting sensitive information from being accessed and captured by unauthorized entities.

To fully satisfy the above requirements, advanced network operation capabilities are required, including *orchestration*, *programmability*, *recursion*, *automated slice management* (self-configuration, self-monitoring, self-optimization, self-healing, etc.), and *exposure*. To exploit them, the operators may rely on the interplay of network softwareization technologies, including virtualization, SDN, and NFV.

### B. Virtualization

Virtualization is key for slicing, as it enables an efficient and secure resource sharing among slices through the use of appropriate abstraction mechanisms. The resources to be virtualized can be physical or already virtualized, supporting a recursive pattern with different abstraction layers.

The introduction of virtualization to networking enables new business models, with novel actors and distinct business roles. We consider a framework with three kinds of actors:

- **Infrastructure Provider (InP):** owns and manages a given physical network and its constituent resources. Such resources, in form of Wide Area Networks (WANs)

and/or data centers (DCs), are virtualized and then offered through programming interfaces to a single or multiple tenants.

- **Tenant:** leases virtual resources from one or more InPs in the form of virtual networks, where the tenant can realize and manage network services. A network service is a logical composition of network functions (NFs). The tenant provide slices as virtual networks with some provisioned network services.
- **End user:** consumes a slice supplied by the tenant, without providing it to other business actors.

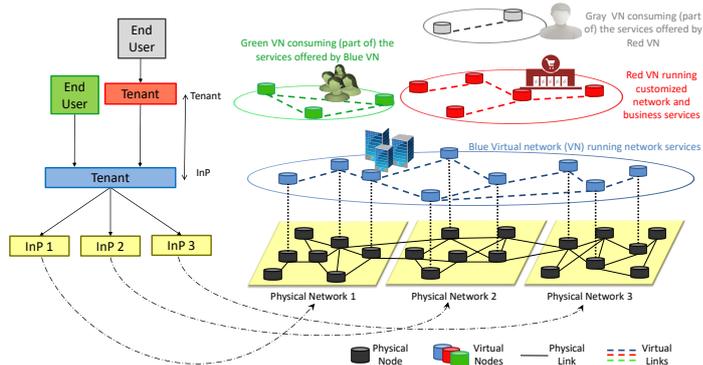


Figure 2. InPs and tenants as main virtualization actors.

As discussed above, virtualization is naturally recursive. Thus, the first two actors can happen in a vertical multi-layered pattern, where a tenant at one layer acts as the InP at the layer immediately above (see Fig. 2).

### C. Software Defined Networking

SDN defines a paradigm where the control plane (policy-based network intelligence) is logically centralized and decoupled from the data plane (packet forwarding/processing).

The SDN paradigm moves the control plane functions from the networking devices to a remote, logically centralized software controller that has a broad view of the devices under its management, and can potentially make better decisions on them. With this approach, devices become inexpensive, programmable forwarding engines that perform data plane functions according to the instructions sent by the SDN controller through a standardized *southbound interface*. This interface enables the controller to real-time program the behavior of the underlying devices. The controller also incorporates a set of *northbound interfaces* through which SDN clients can securely access the controller to program its behavior and gather relevant SDN data for their own management purposes.

SDN flees from traditional time-consuming and error-prone networking practices, setting a environment where network management can be remotely performed in a dynamic, flexible, and programmatic manner.

### D. Network Functions Virtualization

NFV technology seeks for virtualizing traditional NFs (e.g., middleboxes), moving them from proprietary, purpose-built hardware appliances to software applications that can run

on virtual machines (VMs) on top of a commodity, general-purpose hardware infrastructure. In the NFV jargon, these applications are referred to as *Virtualized NFs* (VNFs). To operate VNFs, existing management practices and tools from the cloud computing space can be applied.

With NFV technology, networks become vendor-agnostic platforms where operators can innovate and provision network services with great agility, scalability, and elasticity.

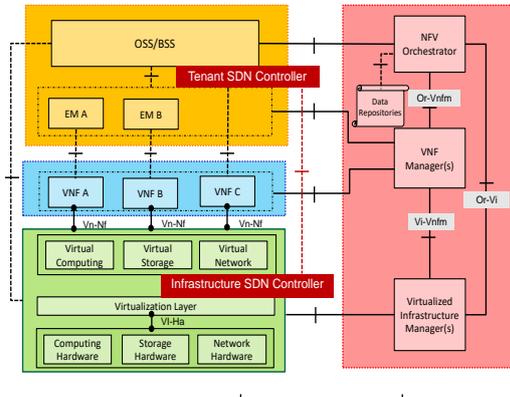


Figure 3. ETSI NFV-MANO architectural framework,

To reduce operational costs in NFV environments, the introduction of automation in network service deployment and management operations is of key importance. To that end, ETSI has defined the NFV Management and Orchestration (NFV-MANO) framework (see Fig. 3). The NFV-MANO framework is broken up into three functional blocks:

- **Virtualized Infrastructure Manager(s) (VIM)**, managing the NFV Infrastructure (NFVI) resources that host and connect VNFs.
- **VNF Manager(s) (VNFM)**, focused on VNF lifecycle management. VNFM interplays with Element Management (EM) blocks, responsible for handling FCAPS management of VNFs.
- **NFV Orchestrator (NFVO)**, logically split into its two main functions: *Resource Orchestrator* (RO) and *Network Service Orchestrator* (NSO). RO orchestrates resources across (potentially different) VIMs, while NSO performs network service lifecycle management.

To incorporate programmability into this framework, ETSI has proposed in [4] the integration of two SDN controllers with distinct roles at different abstraction layers:

- **Tenant controller (TC)**: performs VNF composition to programmatically realize the tenant’s network services.
- **Infrastructure Controller (IC)**: manages NFVI networking resources to programmatically set up VNF connectivity at the infrastructure level (VM connectivity).

### III. 5G NETWORK SLICING ARCHITECTURE

In this section, we describe our architectural proposal for 5G slicing [2], with several network slices running on a common NFVI. The deployment example depicted in Fig. 4 includes two tenants, each managing a particular set of slices. Each slice consists of one or more network services to satisfy the network requirements of the use case it accommodates.

The example considers that the tenant accesses NFVI resources from three InPs. InP1 owns two geographically

remote DCs connected over a shared WAN infrastructure that is composed of two distinct WAN domains. These WAN domains, consisting of SDN transport networking resources that support inter-DC connectivity, are under the control of the two remaining InPs. With this approach, slices can expand across multiple DCs, enabling multi-site NFV deployments.

The VMs and their underlay support (computing and storage resources), are directly managed by a VIM instance at each DC. The networking resources, in search of supporting VM connectivity at both intra-DC (within the DC) and inter-DC (in a WAN domain) level, are programmatically managed by the ICs and WAN ICs, following the VIM and the WAN Infrastructure Manager (WIM) premises, respectively.

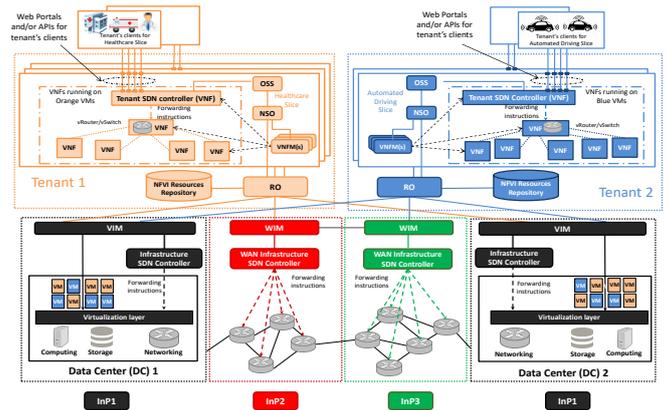


Figure 4. SDN/NFV-based 5G Slicing Architecture.

On top of the InPs, the tenants independently manage a set of network slices. Each slice comprises an OSS, a TC, and an NSO. The OSS instructs the TC to manage slice’s constituent VNFs and logically chain them to efficiently build up the slice’s network service(s). The lifecycle of such network service(s) is managed by the NSO, which interacts with the TC via the OSS. The TC, deployed as a VNF, relies on the capabilities provided by virtual switches/routers to programmatically define the required VNF chaining, forwarding pertinent instructions to such virtual switches/routers via its *southbound interfaces*. Through its *northbound interfaces*, the TC securely exposes network slice capabilities to end users, enabling them to retrieve convenient management data and (if tenant’s and end users’ policies authorizes them) to operate the slice’s constituent network service(s). The fact that each slice has its own NSO, OSS and TC instances enables the required *management isolation*.

Each tenant must efficiently orchestrate their assigned resources to simultaneously satisfy the diverging requirements of the slices that are under its management. The RO is the functional block that performs such task on behalf of the tenant, providing each slice with the required resources via interfaces with each slice’s NSO. The RO must perform the resource sharing among slices while fulfilling their required performance, following an adequate, effective resource management framework that must comply with both tenant and slice-specific policies. This framework is required so that the RO enables *performance isolation* among slices.

All the NFVI resources available for use by a tenant (those that RO orchestrates) are supplied by the different InPs. Each

InP rents part of the virtual resources according to a business lease agreement that both InP and tenant had previously signed. To access, reserve and request such resources, the tenant's RO interacts with the VIM(s) /WIM(s) by means of interfaces that those functional blocks expose and that tenant's RO consumes. Indeed, we assume that VIMs and WIMs support multi-tenancy. We also assume that WIMs can communicate with each other according to predefined business agreements. In this respect, the interaction between a WIM and an RO might be achieved indirectly through another WIM.

As Fig. 4 suggests, the resource management must be performed at two levels: at the infrastructure level, where a slice-agnostic VIM/WIM provides the subscribed tenants with (virtualized) infrastructure resources, and at the tenant level, where the RO delivers its assigned resources to the corresponding slices. Both the VIM(s)/WIM(s) and the RO have to collect accurate resource usage information (each at its domain), and in turn to forecast resource availability in short timescales to satisfy tenant and slice demands, respectively.

To preserve *security and privacy isolation* among slices, it is required to apply the compartmentalization principle at each virtualization level. In addition, each functional block and manageable resource (e.g., VNF) within a given slice must have its own security mechanisms, ensuring operation within expected parameters, and preventing access to unauthorized entities. This is intended to guarantee that faults or attacks occurred in one slice are confined to such slice, preventing their propagation across slice boundaries.

Additionally, note that although recursion has not been addressed in this example, it is readily applicable to this scenario by simply assuming some of the slice's users are tenants which in turn can deploy and operate their own slices.

#### IV. SLICING SUPPORT WITH ETSI OSM STACK

A close collaboration between network engineers, software developers, and operators is essential to ensure a rapid advancement in network slicing. To move from conceptual architectures to slicing-ready testable deployments in an agile and seamless manner, a set of novel software tools (not available today) need to be developed. However, as network slicing strongly relies on SDN and NFV, and there is not a need to reinvent the wheel, current efforts are focused towards how to enhance existing SDN/NFV environments in search of network slicing support. Identifying what capabilities shall be added to these environments is the purpose of this section.

One of the best positioned candidates is OSM, an open-source software stack fully compliant with the NFV-MANO framework. In this section, we first describe the architecture of OSM; then, we use it for a demonstrative slicing-like Proof of Concept (PoC); finally, we identify what OSM misses to get closer to our slicing architecture.

##### A. OSM Architecture

Announced at Mobile World Congress'16, OSM is an ETSI-hosted community that delivers a production-quality open source stack fully aligned with ETSI NFV-MANO framework. With more than 70 partners supporting it (Telefónica, Verizon, Canonical, Intel, etc.), OSM is today the largest and most diverse community in the MANO space.

OSM has a layered architecture (see Fig. 5) consisting of a set of software modules that are compliant with the NFV-MANO functional blocks (see Section II.C). From a bottom-up view, OSM architecture is logically split into the VIM, VCA (VNFM), RO, and NSO layers.

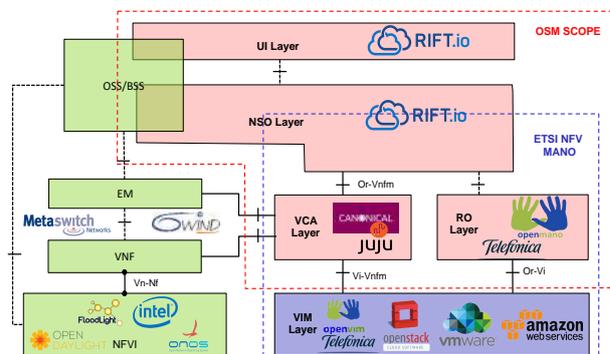


Figure 5. OSM architecture and its constituent software modules.

##### B. A network slicing PoC with OSM

After installing and configuring the different OSM modules shown in Fig. 5, we build up an scenario where two distinct network services run on a common NFVI. Although operating network services is not the same as network slices, this NFV-based deployment can be seen as a slicing-like scenario, just considering that (1) each slice can serve its use case with a single network service, and (2) the two network services have very diverging requirements. With these premises, we use the NSO (Rift.io) to simultaneously launch two network services that concurrently run on shared NFVI, orchestrated and managed by the RO (OpenMANO) and the VIM (OpenVIM), respectively. Network services' constituent VNFs are dynamically managed with the VNFM module (Juju).

##### C. OSM Gaps

After having built up the scenario and explored the OSM capabilities in a profound way, we conclude that although OSM is progressing well, it does not naturally support slicing yet. OSM lacks of a set of fundamental capabilities to be slicing-compliant, including:

- Support for slice management isolation. OSM does not allow a tenant to deploy more than one NSO instance.
- Support for a TC. Thus, a tenant cannot perform VNF chaining in a dynamic, programmatic manner.
- A logical entity that optimally dispatches NFVI resources to slices. OpenMANO lacks this inherent RO functionality in OSM; instead, it acts a simple proxy gateway between the NSO and the VIM layers.
- Support for slice deployments over WAN domains. OSM does not include WIMs nor WAN ICs for a program-mable inter-DC connectivity management.

#### V. RESOURCE ORCHESTRATION FOR 5G SLICING

Resource orchestration is identified as the most challenging issue in network slicing [2]. The RO has to deliver the limited set of resources that are at tenant's disposal in the best way

possible, in search of exploiting multiplexing benefits while satisfying the required performance isolation between slices. The joint fulfillment of both conditions is not an easy task. If the RO decides to share resources among slices, performance isolation is not guaranteed; indeed, if a shared resource fails or suffers from performance degradation, the rest of slices using it may be affected. To overcome this issue, the RO can provide the slices with dedicated resources. However, this strategy leads to multiplexing loss, and may prevent the tenant from profiting from an efficient resource usage.

The above trade-off requires finding a mixed optimal solution that simultaneously exploit the benefits of both approaches. This is the purpose of this section. To that end, we define a scenario where a tenant simultaneously operates three slices, each addressing a particular use case. To exploit the slicing concept, we consider use cases with seemingly disparate performance needs:

- **Gaming (eMBB<sup>1</sup>):** high throughput and spectrum efficiency, and low latency for a seamless user experience.
- **Massive deployment of sensors/actuators (mMTC<sup>2</sup>):** a large set of low-cost, power-constrained devices that transmit low volumes of delay-tolerant data. This use case calls for battery longevity, low maintenance, and extremely high aggregated data rates.
- **Collaborative robots (URLLC<sup>3</sup>):** provides time-critical control of manufacturing and production processes for the industry 4.0. This use case also transmits low payloads; however, unlike mMTC, URLLC demands high reliability and availability, and extremely low latencies.

The structure of this section is as follows: we first derive the traffic models for the use cases; then, we define mathematical frameworks for modelling different RO schemes; finally, we characterize the simulation scenario where we evaluate them.

#### A. Traffic models

Although having disparate performance needs, the mMTC and URLLC use cases share similar traffic patterns, as both rely on Machine-To-Machine (M2M) devices. These devices send small data packets (reports) infrequently to a centralized server in an automated way. This *device-centric* approach, based on the sparse transmission of reports with zero-mobility support for devices, differs from the *user-centric nature* of the eMBB use case, focused on providing a rich user experience through a large set of mobility scenarios. Thus, we define two traffic models: eMBB and M2M models.

For the eMBB model, we adopt the compound data traffic and user behavior model proposed in the scenario described in [5]. This scenario, composed of 12 access points regularly distributed in a 4m x 3m grid over a rectangular area of size 387m x 552m, includes three types of applications (web browsing, Youtube, and Skype services), and considers that users move across the area following a fluid-flow mobility model, with a speed that is uniformly distributed between 0 and 4.2 m/s. The resulting eMBB traffic model is shown in Eq. 1, where  $N_u$  denotes the number of eMBB users,  $\bar{\lambda}_{\text{eMBB}}^u$

the mean arrival rate of traffic packets generated per user, and  $\bar{\lambda}_{\text{eMBB}}$  the aggregated data rate.

$$\bar{\lambda}_{\text{eMBB}} = N_u \cdot \bar{\lambda}_{\text{eMBB}}^u \quad (1)$$

$$\bar{\lambda}_{\text{M2M-x}} = (1 - p_x^{\text{event}}) \cdot \bar{\lambda}_x^{\text{active}} + p_x^{\text{event}} \cdot \bar{\lambda}_x^{\text{event}} \quad (2)$$

For the M2M traffic model, we consider two types of devices. mMTC-type and URLLC-type M2M devices. Additionally, we assume two possible states for each M2M device: *active*, in which the device sends periodical reports to the server, and the *alarm* state, where the device assiduously sends high data volumes (alarms) when an event (e.g., robot malfunctioning in the URLLC) occurs. For both states, we assume that the packet generation is governed by a Poisson process. The M2M traffic model is shown in Eq. 2, where  $\bar{\lambda}_{\text{M2M-x}}$  is the mean arrival packet rate generated by  $N_x$  X-type M2M device, and  $p_x^{\text{event}}$  is the probability that an X-type M2M device to be in the alarm state.

#### B. Resource Orchestration Schemes

In this section, we present three different schemes that may be used to model the way the RO orchestrates the tenant's assigned NFVI resources. These schemes are:

- **Resource sharing:** the resources are shared by the three slices, i.e., all resources simultaneously process the traffic from the three use cases. This scheme maximizes the tenant's profit through an efficient resource usage, but disregards the required performance isolation.
- **Resource dedication:** each slice is provided with dedicated resources. This approach ensures performance isolation at the cost of introducing over-provisioning, particularly when resource delivery strategy considers the need to deliver resources for handling traffic peaks (when a large set of mMTC-type devices are in the alarm state).
- **Traffic shaping:** trade-off between the resource sharing and resource dedication schemes, exploiting multiplexing gains while enabling performance isolation. To achieve this, it adds a traffic shaper that enables a weighted resource sharing among slices, according to a programmable shaping criteria. This criteria may favor the assignment of resources to a given slice to prioritize its traffic processing, and hence fulfilling its performance needs.

To model these schemes, we use the statistical parameters derived from the M/M/m multi-server queuing theory [6].

#### C. Performance Evaluation

In this section, we compute and then compare the performance of the three RO schemes (see Fig. 7. Unlike mMTC, where delay-tolerant data is exchanged, latency is key for eMBB, and even critical for URLLC. For this reason, in this evaluation we set latency  $\bar{T}$  as the reference Key Performance Indicator (KPI) for the performance evaluation. However, in addition to performance behavior, tenants care for the resource costs inherent to each RO scheme. For this reason, and to shed light on the advantages/disadvantages of these RO schemes, we also evaluate and compare the different RO schemes in terms of resource usage (see Fig. 8), mathematically expressed as Service Logics (SLs). From here on out,

<sup>1</sup>enhanced Mobile Broadband is an ITU-T 5G service type

<sup>2</sup>massive Machine-Type Communications is an ITU-T 5G service type

<sup>3</sup>Ultra Reliable and Low Latency Communications is an ITU-T 5G service type

we define a SL as a finite set of NFVI resources (provided by an InP) that a tenant has at its disposal for orchestration.

Simulation Parameters	
Simulation Time	300 s
eMBB users	636,000
M2M mMTC-type devices per eMBB user	9
M2M URLLC-type devices per eMBB user	1
M2M packet size	200 Bytes
M2M event duration	1 s
$\bar{T}_{\max}$	1 ms
Traffic Models	
$\lambda_{eMBB}^u$	0.0588 pkt/s
$\lambda_{active}$	0.0033 pkt/s
$\lambda_{alarm}$	0.033 pkt/s
% M2M mMTC-type devices	90%
Alarm period for mMTC-type devices	60 s
Alarm magnitude values for mMTC-type devices	[30, 50, 80, 50] %
% M2M URLLC-type devices	10%
Alarm period for URLLC-type devices	Unique (at 40 s of the simulation)
Alarm magnitude values for URLLC-type devices	33%

Figure 6. Parameters for the simulation of our slicing-ready scenario.

After generating the traces for eMBB and M2M models with NS3 and MATLAB tools, we build our slicing-ready scenario with the set of parameters summarized in Fig. 6. Getting the scenario ready for simulation, and setting the target latency  $\bar{T}_{\max} = 1$  ms for the eMBB and URLLC slices, we study the different latency values for each RO scheme. Next, we filter the latency values with a simple 90-ms moving average window and compute the Cumulative Distribution Function (CDF), resulting in the performance evaluation shown in Fig. 7.

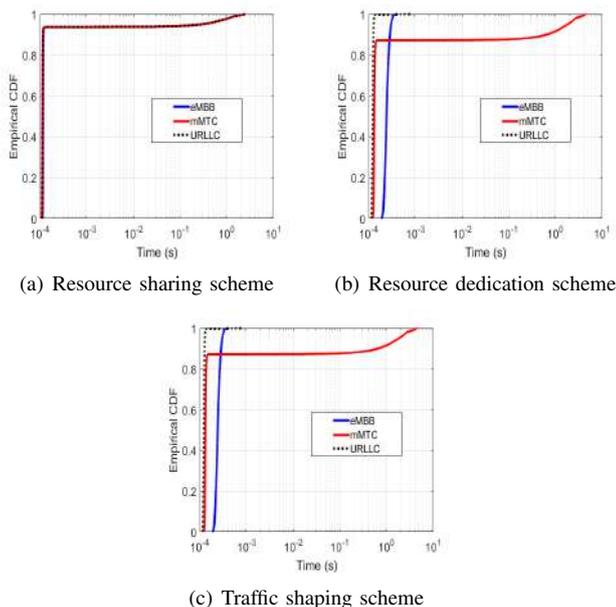


Figure 7. CDF of the filtered system delay for each proposed RO scheme.

The resource sharing scheme minimizes the resource usage by exploiting multiplexing gains (Fig. 8a) at the cost of violating the performance isolation across slices. As seen from Fig. 7a, when mMTC traffic peaks occurs (a percentage of mMTC devices become in the alarm state), not only the performance of the mMTC slice is degraded, but also that of the rest of slices. The fair sharing nature of the scheme is an ill-suited approach in our slicing-ready scenario, as delay-sensitive use cases experience for some time periods latencies significantly higher than  $\bar{T}_{\max}$ .

With the resource dedication scheme, RO prevents performance degradation across slice boundaries at the expense of using the greatest amount of resources (Fig. 8a and b). As seen from Fig. 7b, the eMBB and URLLC slices fulfill their performance targets regardless of what may happen in the mMTC slice, which now suffers a latency increase during a significantly higher time period with respect to resource sharing scheme. However, the delay-tolerant nature of the mMTC use case does not make this latency exceedance so critical as it could be for the rest.

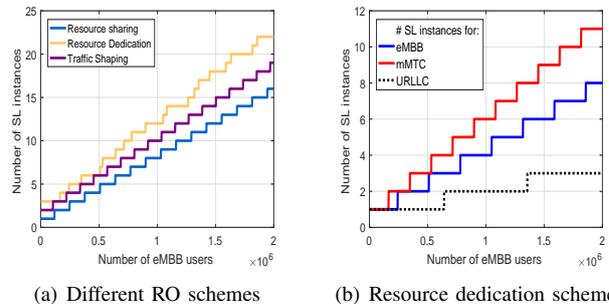


Figure 8. Resource usage for different number of eMBB users and M2M devices (ten M2M devices per user).

Finally, the selected traffic shaping criteria for the last RO scheme (Fig. 7c) improves the results reached above. This is due to the selected traffic shaping criteria, that optimally gives the highest priority to the time-critical use cases. This approach not only improves the performance results, but also enables the tenant to make a more efficient resource usage compared to resource dedication scheme (Fig. 8a).

## VI. CONCLUSIONS

In this work, we have proposed a 5G slicing architecture that capitalizes on the latest advancements in SDN and NFV. Our architecture enables operators to simultaneously serve a large variety of diverging use cases from different vertical industries in a cost-efficient manner. To check the feasibility of this architecture, our solution has been put into context with existing SDN/NFV tools such as those the OSM project uses. Additionally, different schemes for the RO have been proposed and then evaluated in a slicing-ready scenario aligned with our architecture, where a tenant simultaneously operates three distinct slices. The results obtained show the difficulties inherent to the design of an optimal resource orchestration solution, particularly if the objective is to model a scheme able to guarantee performance isolation among slices, and exploit multiplexing gains at the same time.

## REFERENCES

- [1] 5G-PPP, "5G empowering vertical industries", White Paper, Feb. 2016.
- [2] J. Ordóñez-Lucena et al. "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges", *IEEE Communications Magazine*, vol. 55, no. 5, May 2017, pp. 80-87.
- [3] ETSI NFV-EVE 012, "NFV Release 3; Report on Network Slicing Support with ETSI NFV Architecture Framework".
- [4] ETSI GS NFV-EVE 005, "NFV; Ecosystem Report on SDN Usage in NFV Architectural Framework", V1.1.1, Dec. 2015.
- [5] J. Prados-Garzon et al., "Latency evaluation of a virtualized MME", *2016 Wireless Days (WD)*, Toulouse, 2016, pp. 1-3.
- [6] N. Gautan, "Analysis of Queues: Methods and Applications", CRC Press, 2012.

# RADIO RESOURCE ALLOCATION IN PRIVATE 5G NETWORKS

Author: Lorena Chinchilla Romero, e-mail: lorenachinchilla@ugr.es

Tutor: Pablo Muñoz Luengo, e-mail: pabloml@ugr.es

Tutor: Pablo Ameigeiras Gutiérrez, e-mail: pameigeiras@ugr.es

Master: Telecommunication Engineering

Department of Signal Theory, Telematics and Communications of the University of Granada  
University of Granada

**Abstract**—The Fifth Generation of mobile networks is expected not only to enhance mobile broadband and data rates, but also to be capable to support demanding services in terms of latency, reliability and availability. In that respect, 5G is a disruptive technology that allows the support of any kind of service. This feature makes 5G a very appealing technology for private networks. The radio access network of private 5G networks are often characterized because of being heterogeneous networks with multiple access technologies (WiFi, 5G), in order to cope with the requirements and specifications of different types of services that need to be met in these scenarios. In this work, a radio resource allocation algorithm is developed for a multi-technology RAN of an industrial scenario in which two kind of services in terms of networks requirements are provisioned.

**Index Terms**—5G, private network, WiFi, multi-technology RAN, network slicing, Industry 4.0.

## I. INTRODUCTION

THE Fifth Generation (5G) of mobile networks is envisioned to accommodate a broad range of use cases requiring different type of services and new business models in an agile and cost efficient manner. To that end, network slicing and multi-tenancy are technological key enablers in 5G systems. With network slicing the physical mobile network infrastructure can be shared among different communication service providers. Furthermore, 5G network slices can be adapted to the variety of needs of each specific vertical industry in a dynamic way. On the other hand, with multi-tenancy paradigm a third-party will make use of the shared infrastructure as a tenant.

An emerging use case in 5G, thanks to the paradigms aforementioned, is the appearance of private networks, also called Non-Public Networks (NPNs), due to the demanding and critical communications of vertical use cases. A report made by Mobile Experts forecasts a significant growth of the private Long Term Evolution (LTE) and 5G markets, resulting in about 3.4 billions of dollars of revenue by 2024 [1]. 5G is a disruptive technology that supports any kind of services, what makes it attractive and suitable for private networks.

Industry 4.0 is arising as a relevant use case of this kind of private networks. Many of these environments have already deployed other radio access technologies such as Wi-Fi, but this is neither enough nor suitable for the supporting of critical

services. For that reason, these networks will harness the technologies they have already deployed in addition to the introduction of 5G technology. So that, the Radio Access Network (RAN) of private 5G networks are characterized by having multiple radio access technologies. Moreover, the resource provisioning task in these networks is particularly adapted to the vertical necessities, what differs them significantly from the public networks.

Conversely, spectrum is a key asset to run a private network as radio resources are an expensive and scarce good. So that, correct radio resource allocation and management are crucial mechanisms to leverage the capabilities of the used wireless access technologies, and in that way get the optimum network performance.

All in all, the emergence and relevance of private networks deployments in industrial scenarios, their particularity of being characterized by multi-Wireless Access Technology (WAT) RANs, and how to manage the set of radio resources in order to make an efficient resource allocation to different nature network slices with the aim of satisfying their corresponding service demands and requirements are the main motivation that drives the realization of this project. Thus, the main objective of this work is to address the radio resource allocation problem in a private 5G network, specifically in an industry 4.0 characterized by multiple radio access technologies, in order to satisfy the traffic demands of services of different natures.

## II. STATE OF THE ART

The radio resource allocation problem in wireless networks has already been tackled in the literature. The work in [2] addresses the problem of resource allocation and packet scheduling for a mixture of Ultra-Reliable and Low Latency Communications (URLLC) and enhanced Mobile BroadBand (eMBB) traffic in 5G networks. In this work, a novel resource allocation method that is latency, control channel, hybrid automatic repeat request and radio channel aware is presented. Furthermore, several works have developed radio resource allocation algorithms for LTE networks. The work in [3] proposes a downlink resource allocation algorithm that uses Quality of Service (QoS) requirements and channel conditions

to address performance degradation. This algorithm is formulated as an optimisation problem where network resources are allocated according to users' priority, whereas the scheduling algorithm decides on the basis of users' channel status to satisfy the demands of QoS.

With respect to Wi-Fi networks, there are also some works that address the resource allocation problem in this technology. In [4], RAN slicing in Wi-Fi networks is addressed. A scheduling algorithm that allocates airtime to a set of virtual interfaces executing on the same or on different physical radios is developed. The work in [5] presents a novel virtualized Wi-Fi network hypervisor based on a time variant radio resource sharing mechanism named Weighted Air-Time Deficit Round Robin (WADRR) that is able to follow the dynamicity of the traffic variations seen by the different tenants located to the Access Points (APs).

Cellular networks with heterogeneous Radio Access Technology (RAT)s in which Wi-Fi coexists with 3G/4G have been studied for many years in order to offload the overloaded technology and so that, make an enhancement in the network performance. In [6] a load balancing scheme for the LTE cellular and Wi-Fi coexisted networks is proposed. The work in [7] proposes a load-balancing scheme for a cellular-Wireless Local Area Network (WLAN) heterogeneous network, where the user association is controlled by employing a cell-breathing technique for the WLAN network.

Since private 5G networks and their uses cases like Industry 4.0 are emerging at the time of writing this project, not many works have been addressed regarding the resource allocation problem.

In spite of the fact that there exist abundant literature, as it can be realized the problem to be solved appears fragmented (resource allocation, private scenarios, multi-technology RANs, network slicing, and son on). This is the reason why this work investigates these issues together.

### III. SYSTEM DESIGN

This section gives a general description of the network scenario design. Moreover, the problem formulation and all the models and assumptions considered are set out.

#### A. System model

As it was stated in the introduction, this project puts its focus on a Non-Public Network (NPN) scenario, specifically in an Industry 4.0 environment. An important particularity of the network scenario considered is the heterogeneity of the RAN. This RAN is specifically characterized by having multiple wireless access technologies, which are 5G New Radio (5G NR) (Third Generation Partnership Project (3GPP) access type) and Wi-Fi (non-3GPP access type). On the top of the mentioned wireless network infrastructure, two different tenants or slices providing two different kind of services are deployed. It is worth clarifying that this problem could be seen in two different ways: there could be two tenants providing each of them a different service on the top of the

same network infrastructure; or just one tenant providing two different services.

On the one hand, in our particular scenario, one of the slices which is referred to as an eMBB slice provides service to user-centric applications requesting high traffic capacity requirements (e.g. virtual reality applications). On the other hand, the second type of slice deployed on the top of the RAN is referred to as an URLLC slice with the purpose of providing service to the most critical applications in terms of latency (e.g. mobile robots like Automated Guided Vehicles (AGVs) helping in assets transport tasks).

Fig. 1 represents the concept of the deployed network scenario. The layer at the bottom constitutes the multi-technology radio access network with 5G NR and Wi-Fi as the wireless access technologies. Observe that the cells representing the base stations coverage area are coloured in two different tones (orange and blue). This detail represents that 5G Base Stations (BSs) will be able to serve users belonging to both eMBB and URLLC slices. In contrast, Wi-Fi access points will be able to serve just eMBB users. The other two layers above represent the two aforementioned types of slices. The one in orange colour represents different services that can be classified as 5G eMBB services. Lastly, the layer on top of the figure (in blue colour) depicts some services or applications, which in this case can be classified as URLLC 5G services.



Fig. 1. Private 5G network scenario.

For the development of this project it is considered a system-level simulator. This simulator contemplates a multi-WAT RAN with two different coexisting technologies (5G NR and Wi-Fi, as it has been mentioned). User Equipments (UEs) are distributed in the scenario taking a fixed position. In this scenario, we consider two kind of users: users that demand services of the eMBB slice, or the ones that demand services of the URLLC slice. Depending on the service the users request, they can access the network by means of a BS or Wi-Fi AP. In this work we are not considering that Wi-Fi technology is aware of QoS. So that, the stringent characteristics of URLLC traffic in terms of latency and availability make Wi-Fi not be a suitable access technology to drive this kind of traffic. Conversely, the traffic classified as eMBB is characterized by higher data rate requirements. In order to offload 5G RAN, Wi-Fi technology will be treated as the priority access for eMBB applications.

No interference is assumed between BSs and APs since different spectrum bandwidths are used by each of the technologies. However, inter-cell interference is considered among BSs, and interferences are also considered among Wi-Fi APs. Regarding the resource allocation task in each deployed technology:

- In **5GNR BSs**, resource allocation is performed at channel level. A channel can be seen as a chunk or set of Physical Resource Blocks (PRBs), or equivalently as a 3GPP Resource Block Group (RBG). In 3GPP TS 36.213 [8], it is specified that the resource allocation task can be done at the level of RBG (i.e., the resource blocks are grouped in several RBGs).
- In **Wi-Fi APs**, the three non overlapping channels (see (a) of Fig. 2) can be allocated to them (i.e., each of the channels 1, 6 and 11 could be allocated to a different AP). Moreover, channel bonding technique is also used when the AP is located in a certain point with high traffic demand. So, in this case, the secondary channels that are contiguous to the primary ones are also allocated to these specific APs (see (b) of Fig. 2). In every AP, the type of resource that can be allocated to each slice is a fraction of the total available airtime.

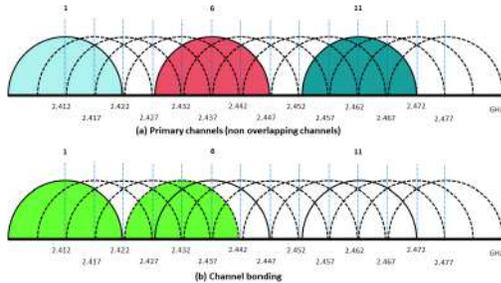


Fig. 2. Primary channels and channel bonding in APs.

## B. Problem Formulation

The objective of the problem addressed is twofold: (i) the maximization of the number of users that can be served at the same time as the service performance requirements are met considering the radio resources and bandwidth limitations (i.e., to minimize the number of rejected users), and (ii) the maximization of Wi-Fi offloading.

The problem is subject to a set of constraints:

- A UE will be associated with a BS or a Wi-Fi AP as long as the Signal to Interference plus Noise Ratio (SINR) that the user perceives from a certain point of attachment is higher than  $-5$  dB, otherwise the user will be directly rejected:

$$SINR_{i,j} > -5dB \quad (1)$$

- The delay requirement for a URLLC user has to be smaller or equal to the specific application delay constraint requested by that user:

$$D_j^u \leq D, \forall j \in N^u \quad (2)$$

- The number of users served by an AP must not exceed a maximum  $N_{max}$ . Otherwise, Wi-Fi APs are considered to be overloaded:

$$\sum_{j \in N^e} N_{j,k}^e \leq N_{max}, \forall k \in N_{APs} \quad (3)$$

- UEs must be guaranteed a specific bitrate, which differs depending on the type of service they request:

$$Th_j^u \geq GBR_u, \forall j \in N^u; \quad (4)$$

$$Th_j^e \geq GBR_e, \forall j \in N^e \quad (5)$$

- The bandwidth resources allocated to the two slices have to satisfy the following expression:

$$\sum_{s \in S^e} BW^e + \sum_{s \in S^u} BW^u \leq BW_{system} \quad (6)$$

This expression represents that the total bandwidth allocated to both type of slices (eMBB and URLLC) must be smaller or equal to the available system (i.e. the total available bandwidth in the network scenario considering 5GNR and Wi-Fi nodes).

To solve this problem we need some analytical models that provide us the equations from which we can obtain the users' required bandwidth.

## C. Performance Modelling

This section gathers all the models that somehow represent some characteristics of the network scenario considered.

On the one hand, traffic of UEs belonging to applications that request eMBB services is characterized by a constant Guaranteed Bit Rate (GBR):

$$GBR_e = C \quad (7)$$

where  $C$  represents a constant value of the eMBB users guaranteed bitrate.

On the other hand, the traffic of UEs that belong to applications that request reliable and low latency communications has to be somehow characterized by the delay constraint required by the particular application. Hence, to characterize this URLLC traffic we have based on the model of work [9]. From this work we obtain the following expression:

$$GBR_u = \frac{\sigma}{D + \frac{\sigma}{P}} \quad (8)$$

where  $\sigma$ ,  $P$  and  $D$  represent the maximum packet size, peak data rate and service delay requirement, respectively. With this expression we assign a user a GBR that is smaller than its required peak data rate, but enough to satisfy the delay requirements, also considering the packets size.

With the focus on calculating the bandwidth demanded by each kind of user in our scenario, it is necessary to obtain this information from analytical models in order to be confident that the UEs QoS requirements are guaranteed. According to our system model, the amount of bandwidth requested by a

user depends on the access technology (i.e. 5G NR or Wi-Fi) and on the type of service delivered by 5G technology (i.e. whether the slice serving the UE is eMBB or URLLC type). Based on the works in [10] and [11], the achievable rate of the users served by an eMBB slice can be captured by Shannon's capacity in 3GPP Single-Input Single-Output (SISO) systems, as expressed below:

$$R_j^e = BW_j^e \cdot \log_2(1 + SINR_{i,j}) \quad (9)$$

where  $BW_j^e$  denotes the bandwidth required by user  $j$  in an slice,  $SINR_{i,j}$  stands for the perceived SINR by user  $j$  attached to base station  $i$ . However, this expression cannot be used for an URLLC slice. This kind of communications are characterized by very low latency applications with small packet sizes, increasing the probability of dropped transmissions due to noise effects. Thus, data must be encoded at a significant lower rate than the Shannon capacity in order to get higher reliability. For these reasons, the capacity achieved by URLLC users served by 5G NR derived in the works [10] and [11] has the following expression:

$$R_j^u = BW_j^u \cdot \left( \log_2(1 + SINR_{i,j}) - \sqrt{\frac{C_{j,s}}{n}} \cdot Q^{-1}(\epsilon) \cdot \log_2(e) \right) \quad (10)$$

where:

- $BW_j^u$  is the bandwidth required by user  $j$  in URLLC slice.
- $SINR_{i,j}$  is the perceived SINR by the user  $j$  when it is attached to base station  $i$ .
- $n = 168$  symbols, which is inspired by the size of a resource block in an LTE system, that contains  $12 \times 7 = 84$  symbols and lasts 0.5 ms [11].
- $C_{j,s}$  is the channel dispersion of UE  $j$  in the slice  $s$ . Its value is given by the following expression:

$$C_{j,s} = 1 - \frac{1}{(1 + SINR_{i,j})^2} \quad (11)$$

- $Q^{-1}(\cdot)$  is the inverse of the Gaussian Q function.
- $\epsilon$  stands for the transmission error probability. Its value is greater than 0.
- $\log_2(e)$  refers to the logarithm in base 2 of number  $e$ .

Observe that expression (10) adds a correction factor to the Shannon's capacity formula in order to consider the specific physical layer behaviour for ultra-reliable and low latency communications with small packet sizes.

Conversely, in the calculus of the capacity achieved by eMBB users served by Wi-Fi two parameters representing the physical and MAC layer effects are considered due to the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol in which this technology relies on. So that, based on the models of works [12] and [13] we propose the expression below in which  $T_i^{mac}$  stands for the interference effects and the load of AP  $i$ , and  $B$  represents the bandwidth.

$$\beta_{i,j}^w = B \cdot \log_2 \cdot \left( 1 + \frac{SINR_{i,j}}{\eta_{phy}} \right) \cdot T_i^{mac} \quad (12)$$

From these three models (9, 10 and 12), we can directly derive the bandwidth each user requires. Note that it depends on the SINR, which in turn, varies according to the interferences. This fact makes the problem difficult to be solved as an optimization problem. For this reason, a heuristic-based algorithm has been developed.

#### D. Radio resource allocation algorithm

The main purpose of the proposed algorithm is the radio resource allocation to users demanding two kind of services in a multi-technology radio access network. Its objective is two fold: minimize the number of rejected users and maximize the number of eMBB users offloaded to Wi-Fi. Fig.3 depicts the algorithm overview with a flow diagram.

The two main ideas that summarize the algorithm procedure are as follows:

- 1) URLLC users are assumed to be served only by 5G technology, so that they have priority to be served by a BS (based on the SINR level and bandwidth availability).
- 2) eMBB users will be served through Wi-Fi technology preferently in order to offload 5G from these users as much as possible in such a way that BSs are available for URLLC users. If any Wi-Fi AP couldn't serve some of these users, the algorithm will try to associate them with a BS whether they can be guaranteed a minimum SINR level and the bandwidth required.

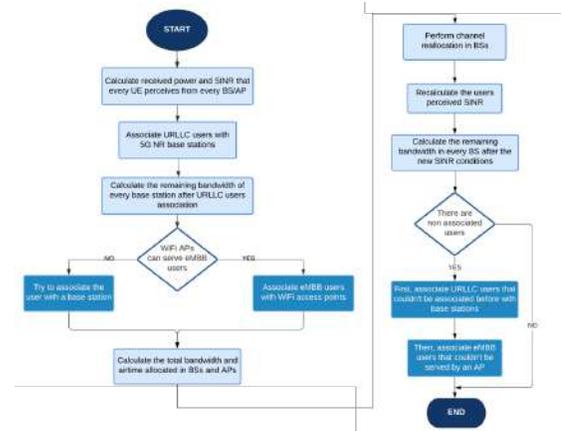


Fig. 3. Radio resource allocation algorithm overview.

#### IV. EXPERIMENTAL SETUP

Figure 4 sketches the simulation scenario. The RAN is shared among two coexisting network slices, one of them tailored for user-centric applications that require high data rates (i.e., eMBB applications) and the other for URLLC services. The GBR characterizing the traffic sources is enforced at the RAN part dedicating bandwidth resources for the user. For the sake of simplicity, it was assumed that all eMBB and URLLC traffic sources have the same GBR, being the eMBB GBR different from the URLLC one.

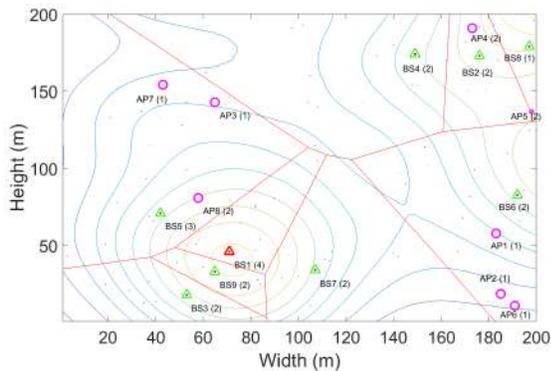


Fig. 4. Realization of the simulation scenario.

The contour lines in Fig. 4 represent the density of UEs, which obeys a Gaussian distribution. It is worthy of mention that the planning of the RAN was performed taking into account the spatial distribution of the traffic demand. By way of illustration, observe that the micro base station is placed close to the center of a traffic hotspot.

### V. RESULTS AND PERFORMANCE ANALYSIS

In this section we provide the analysis of the most significant results.

Figure 5 depicts the Cumulative Distribution Function (CDF) of the measured spectral efficiency for both eMBB UEs and URLLC UEs. As observed, overall, URLLC users achieve higher spectral efficiencies than eMBB ones. This fact is mainly due to the URLLC users are only served by 5G, which has higher spectral efficiencies than Wi-Fi, as enforced by the proposed algorithm in this project.

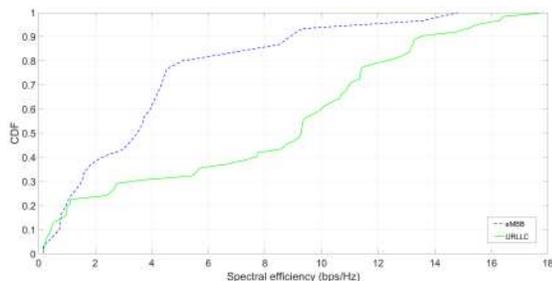


Fig. 5. CDF of the demanded spectral efficiency per slice.

Figure 6 includes the empirical CDF of the bandwidth demanded by each type of UE. The results show that eMBB users require a larger amount of spectrum resource than URLLC ones. This result is due to eMBB users are preferently served by Wi-Fi, which has smaller spectral efficiencies, and also because a greater GBR is considered for eMBB users in this scenario (5 Mbps compared to 2.43 Mbps of URLLC users).

Figure 7 depicts an isolated experiment that compares the demanded bandwidth by both technologies under the same SINR conditions and the same GBR to be guaranteed. Wi-Fi technology requires higher bandwidths than 5G to enforce

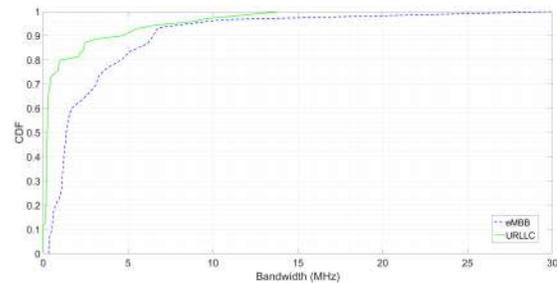


Fig. 6. CDF of the demanded bandwidth per slice.

a given GBR for a given user SINR value. We have already pointed out that Wi-Fi has lower spectral efficiencies than 5G, which is the explanation of this result.

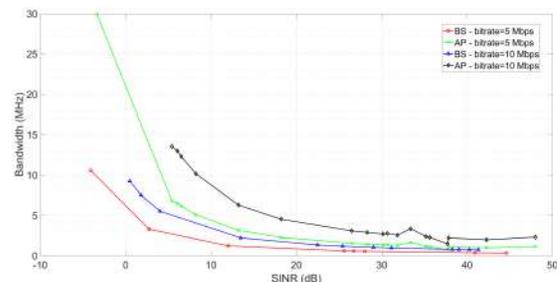


Fig. 7. BW versus SINR required by eMBB users in APs and BSs for two different values of bitrate.

Figure 8 represents the impact the SINR has on the bandwidth demand of the URLLC UEs for different latency constraints. As expected, the more stringent the delay requisite is, the higher the bandwidth demanded will be. Observe that the difference in the bandwidth demand for the two values of the URLLC delay budgets considered significantly increases for poor SINRs. These results suggest that a careful RAN planning is of utmost importance in scenarios with ultra-reliable and low latency communications in order to keep the radio resource demand of these UEs within a reasonable and acceptable margin.

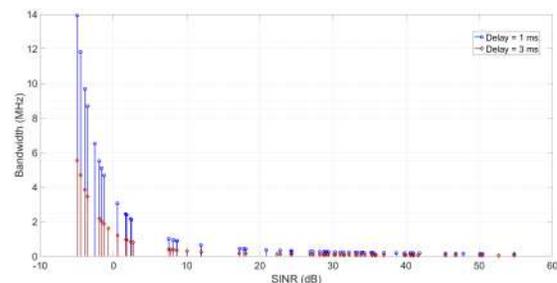


Fig. 8. BW versus SINR for two values of the URLLC delay budget.

The same occurrence can be seen indirectly in Fig. 9 where the number of rejected users against the delay budget of URLLC users is represented. Note that the number of rejected users is higher when more stringent is the delay requirement.

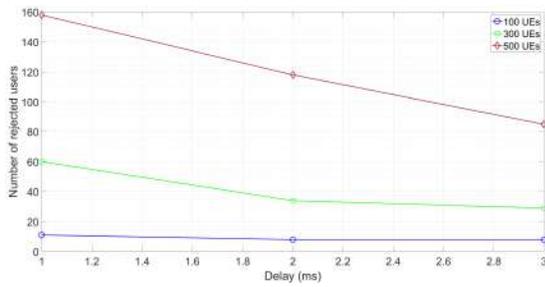


Fig. 9. Number of rejected users versus the values of URLLC delay budget.

Lastly, Fig. 10 shows the number of eMBB users offloaded to Wi-Fi versus the total number of users simulated in the scenario. It has to be highlighted that these results depends on the considered scenario (i.e. RAN planning, users spatial distribution, traffic demand). It can be seen that over 55 users can be served by Wi-Fi. Taking into account that the number of eMBB users in the scenario conforms 30% over the total represented, the results show that Wi-Fi offloads a significative amount of users.

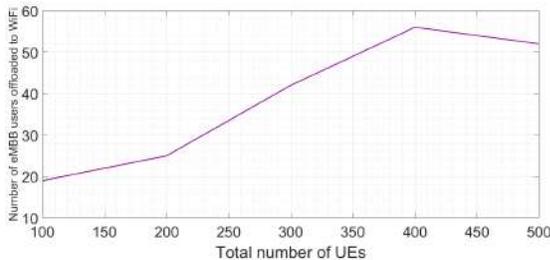


Fig. 10. Number of eMBB users offloaded to Wi-Fi versus the total number of users in the scenario.

## VI. CONCLUSIONS

From the results obtained, several statements can be derived. They are indicated as follows:

- The proposed algorithm proves to be an effective approach on serving URLLC traffic and on offloading eMBB traffic to Wi-Fi.
- 5G radio access technology provides higher spectral efficiencies than Wi-Fi. This makes 5G more suitable for critical or demanding applications in dense network scenarios in case only one RAT is deployed.
- In spite of the fact that Wi-Fi is not suitable to support URLLC traffic, it is an appealing technology to offload 5G network from eMBB traffic, and so that decrease deployment and operation costs.
- In this kind of environments the points of attachment location is crucial when putting the focus on making an efficient use of radio resources while the number of rejected users is desired to be as minimum as possible.

## ACKNOWLEDGMENTS

I would like to express my deepest appreciation to all the people that in some way have been important in getting

this Master's Thesis to its completion. Thanks to WiMuNet Lab group for letting me be a member of the research group, and specially to my supervisors Pablo Muñoz and Pablo Ameigeiras for their support, ideas and guidance. I heartily thank my parents, brother and sister for their tireless effort, their support and because despite the distance and the circumstances I always feel them close. Thanks to Alberto for his unfailing support and words of encouragement. I would also thank my research colleagues and friends for giving me moments of comfort and relief.

## REFERENCES

- [1] "Mobile Experts forecast," available at <https://www.prnewswire.com/news-releases/mobile-experts-inc-releases-industry-benchmark-report-on-private-lte-300793642.html> (last accessed on 21/10/2020).
- [2] A. Karimi, K. I. Pedersen, N. H. Mahmood, G. Pocovi and P. Mogensen, "Efficient Low Complexity Packet Scheduling Algorithm for Mixed URLLC and eMBB Traffic in 5G," 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), Kuala Lumpur, Malaysia, 2019, pp. 1-6, doi: 10.1109/VTCSpring.2019.8746407.
- [3] M. Mamman, Z. Hanapi, A. Abdullah, A. Muhammed, "Quality of Service Class Identifier (QCI) radio resource allocation algorithm for LTE downlink," in PLOS ONE, vol. 14, pp. 1-22, January 2019, doi: 10.1371/journal.pone.0210310
- [4] J. J. Aleixendri, A. Betzler and D. Camps-Mur, "A practical approach to slicing Wi-Fi RANs in future 5G networks," 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 2019, pp. 1-6, doi: 10.1109/WCNC.2019.8885777.
- [5] K. Koutlia, A. Umbert, R. Riggio, I. Vilà and F. Casadevall, "A new RAN slicing strategy for multi-tenancy support in a WLAN scenario," 2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft), Montreal, QC, 2018, pp. 64-70, doi: 10.1109/NETSOFT.2018.8460138.
- [6] Zunli Yang, Qinghai Yang, Fenglin Fu and Kyung Sup Kwak, "A novel load balancing scheme in LTE and WiFi coexisted network for OFDMA system," 2013 International Conference on Wireless Communications and Signal Processing, Hangzhou, 2013, pp. 1-5, doi: 10.1109/WCSP.2013.6677253.
- [7] S. Navaratnarajah, M. Dianati and M. A. Imran, "A Novel Load-Balancing Scheme for Cellular-WLAN Heterogeneous Systems With a Cell-Breathing Technique," in IEEE Systems Journal, vol. 12, no. 3, pp. 2094-2105, Sept. 2018, doi: 10.1109/JSYST.2017.2733446.
- [8] 3GPP TS 36.213 Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Rel-16), V16.2.0 Jun. 2020.
- [9] H. Alshaer and H. Haas, "SDN-enabled Li-Fi/Wi-Fi wireless medium access technologies integration framework," 2016 IEEE Conference on Standards for Communications and Networking (CSCN), Berlin, 2016, pp. 1-6, doi: 10.1109/CSCN.2016.7784888.
- [10] J. Tang, B. Shim and T. Q. S. Quek, "Service Multiplexing and Revenue Maximization in Sliced C-RAN Incorporated With URLLC and Multicast eMBB," in IEEE Journal on Selected Areas in Communications, vol. 37, no. 4, pp. 881-895, April 2019, doi: 10.1109/JSAC.2019.2898745.
- [11] Sebastian Schiessl and James Gross, "Delay analysis for wireless fading channels with finite blocklength channel coding," Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, pp. 13-22, 2015.
- [12] S. Navaratnarajah, M. Dianati and M. A. Imran, "A Novel Load-Balancing Scheme for Cellular-WLAN Heterogeneous Systems With a Cell-Breathing Technique," in IEEE Systems Journal, vol. 12, no. 3, pp. 2094-2105, Sept. 2018, doi: 10.1109/JSYST.2017.2733446.
- [13] L. Simić, M. Petrova and P. Mähönen, "Wi-Fi, but not on Steroids: Performance analysis of a Wi-Fi-like Network operating in TVWS under realistic conditions," 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, 2012, pp. 1533-1538, doi: 10.1109/ICC.2012.6364296.

# SDN and NFV in 5G Networks: Handover and Automated Network Service Scaling as Use Cases

Author: Óscar Adamuz Hinojosa, e-mail: oadamuz@ugr.es

Supervisor: Pablo Ameigeiras Gutiérrez, e-mail: pameigeiras@ugr.es

Supervisor: Juan José Ramos Muñoz, e-mail: jjramos@ugr.es

Master's Degree: Telecommunication Engineering

Department of Signal Theory, Telematics and Communications

University of Granada

**Abstract**—SDN and NFV are considered 5G key enablers. In this work, they are applied to 5G network procedures such as handover and automated network service scaling. On the one hand, an SDN-based handover in a partially virtualized LTE architecture is designed. It is characterized by the substitution of GTP protocol for MPLS, which reduces packet overhead. Furthermore, a simulator is developed to verify its performance. It implements the messages exchange and the EPC entities processing. On the other hand, automated network service scaling procedures are proposed. They involve an optimum resource usage while face the dynamic traffic variation. Moreover, NFV templates are also applied to describe a virtualized EPC (vEPC). Additionally, theoretical and simulated scenarios are presented. The former describes a vEPC scaling, whereas the latter evaluate a dynamic resource provisioning algorithm used for scaling a virtualized MME (vMME).

**Index Terms**—5G Networks, SDN, NFV, Handover, Scaling, VNF, Network Service, EPC, MME.

## I. INTRODUCTION

THE increment of mobile devices along with the appearance of new services will cause an exponential growth of data traffic and the emergence of challenging requirements. Thus, 4G network would be drastically upgraded in such a way that a new mobile network generation will really result [1]. To fulfill 5G requirements, some technologies as Software Defined Networking (SDN) and Network Function Virtualization (NFV) have been proposed by researchers. These technologies bring operators the promise of reducing CAPEX and OPEX [2], enhancing network flexibility and scalability, and shortening the time-to-market of new services.

In this work, these technologies are applied to two relevant 5G scenarios: handover and automated network service scaling. The former scenario involves user mobility between eNBs. It is designed in a partial virtualized LTE architecture, where SDN plays a key role switching the forwarding paths between eNBs and packet data network. The analyzed impacts of SDN and NFV in this procedure along with latency measurement are contribution objects in a conference paper [3]. The last scenario allows an automatic adaptation of the amount of virtualized resources required by Network Services (NSs). It is achieved through scaling procedures, whose workflows are compliance with ETSI specifications. These workflows are also gathered in a paper which is currently under review in IEEE Network Magazine.

The remainder of this document<sup>1</sup> is organized as follow. First, an overview of SDN, NFV and their relationship is provided in Section II. Then, a SDN and NFV-based architecture is specified in Section III. Next, handover and automated NS scaling procedures and their implications are presented in Section IV. Later, they are evaluated in Section V. Finally, conclusions are remarked in Section VI.

## II. BACKGROUND

In this section, main concepts of SDN, NFV and their relationships are overviewed.

### A. Software Defined Networking (SDN)

SDN is a technology which enables network programmability through a centralized control logic. SDN networks are mainly characterized by the control and data plane separation, in contrast with traditional networks whose planes are vertically integrated. Data plane is composed of packet-forwarding devices, whereas control plane is constituted by a SDN controller with a global network vision. Furthermore, SDN presents a management plane which consist of SDN applications [5].

The SDN controller interacts with data and management plane through southbound and northbound APIs. The southbound API is used to change forwarding rules in data plane devices. The most well-know protocol accepted is OpenFlow [6]. The northbound API acts as interface between applications and SDN Controller. Currently there are not an unique and accepted protocol.

### B. Network Function Virtualization (NFV)

NFV is a technology which decouples software applications from specific hardware, and runs them in generic high volume servers located in datacenter infrastructures [7]. This paradigm enables sharing infrastructure resources, thus operators can deploy new NSs faster, providing greater flexibility to scale them dynamically.

Each NS [8] is composed of Virtualized Network Functions (VNFs) as shows Fig. 1. Virtual Links (VLs) are abstractions of physical links that logically connect together VNFs.

<sup>1</sup>This document is a merely self-contained summary of the Master's Thesis. For further details consult [4]

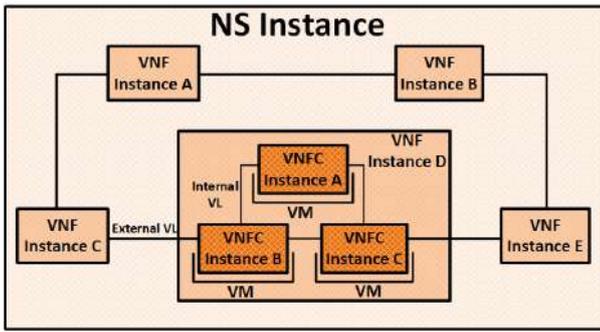


Figure 1: Network Service internal composition

To specify how VNFs are connected, each NS may have one or more VNF Forwarding Graphs (VNFFGs). A VNFFG defines the topology of the NS, and optionally one or more Network Forwarding Paths (NFPs). These NFPs include forwarding rules to appropriately steer the traffic over this topology.

For a fine-grained control of its performance and reliability, a VNF might be decomposed into one or more VNF Components (VNFCs), each performing a well-defined subset of the entire VNF functionality. Each VNFC is hosted in a single virtualization container.

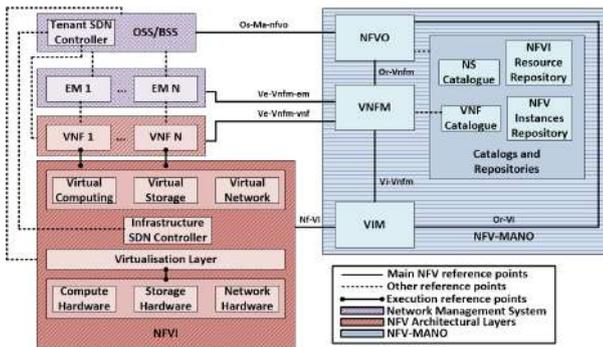


Figure 2: Integration of Infrastructure and Tenant SDN Controllers into the ETSI NFV Architectural Framework. Adapted from [8] [9] [10]

To automate the deployment and management of NSs, ETSI defines the NFV Management and Orchestration (NFV-MANO) architecture [11]. The NFV-MANO stack is broken up into three functional blocks as shows Fig. 2

- Virtualized Infrastructure Manager(s) (VIM), managing the NFV Infrastructure (NFVI) resources that host and connect VNFs.
- VNF Manager(s) (VNFM), focused on VNF lifecycle management. VNFM interplays with Element Management (EM) blocks, responsible for the Fault, Configuration, Accounting, Performance, Security (FCAPS) management of VNFs.
- NFV Orchestrator (NFVO), that orchestrates NFVI resources across VIMs and performs NS lifecycle management. NFVO may interact with traditional Operations/Business Support Systems (OSS/BSS).

To incorporate programmability into the NFV architecture, the integration of two SDN controllers with distinct roles is proposed in [9]:

- Tenant controller: Part of the OSS or deployed as a VNF, performs VNF composition to dynamically realize NSs.

- Infrastructure controller: Managed by the VIM, implements VNF connectivity at the NFVI.

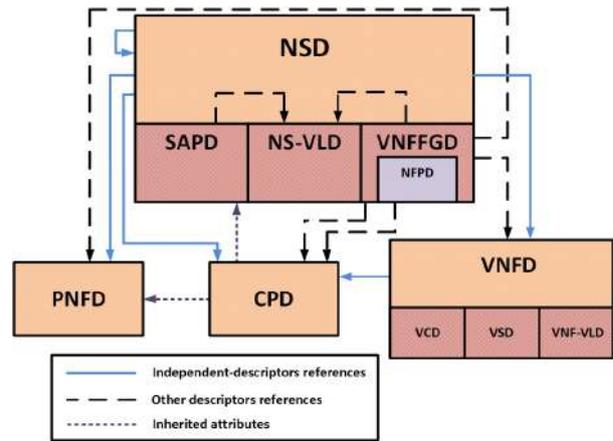


Figure 3: Network Service Descriptor Structure

The NFV-MANO blocks use descriptors (see Fig. 3) that contains machine-processable information to create instances of an NS, and manage them throughout their lifetime. The main descriptors are NS Descriptor (NSD) [12] and VNF Descriptor (VNF) [13]. Both are composed by other descriptors and attributes which are interrelated.

### C. SDN and NFV relationships

Although SDN and NFV might be applied independently, they can cooperate to increase their accumulate value [14]. An example of this cooperation is shown in Fig. 2), in which NFV provides the framework that orchestrates and manages the NSs, whereas SDN controllers incorporates the connectivity between NS constituents.

## III. 5G ARCHITECTURE BASED ON SDN AND NFV

In this section, a state-of-the-art review is realized to design the system architecture.

### A. SDN and NFV Integration in EPC Architecture

Evolved Packet Core (EPC) presents a set of drawbacks which preclude the performance improvement required to achieve 5G requirements. To resolve them, EPC might make use of SDN and NFV. Therefore, researchers have proved the feasibility of SDN and NFV integration into EPC in their works. Among these, it highlights those related with legacy EPC migration to SDN and NFV-based EPC. In [15] and [16], the authors present different options to integrate SDN controllers, and virtualize EPC entities respectively. In [17], the authors give 3 phases to progressively realize this migration. At first, the EPC entities are virtualized. Then, SDN paradigm is included, while still maintaining legacy nodes. Finally, the network is fully SDN-compliant, and GTP tunneling and legacy nodes at the data plane are removed.

### B. System Architecture

The principles given previously along with the architecture proposed in [18] are used for designing the system architecture which will support handover and automated NS scaling procedures. It is depicted in Fig. 4, and consist of three planes:

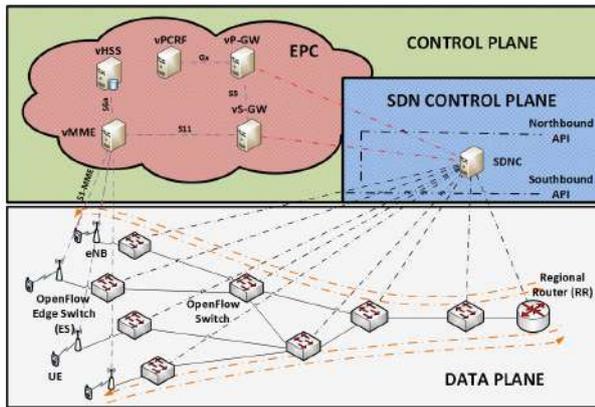


Figure 4: System Architecture Design

- SDN Control Plane: Contains a SDN Controller integrated with S-GW, which allows a complete separation between control and data plane functionalities.
- Control Plane: Comprise EPC control functionalities. They are implemented as VNFs following an 1:N option (descomposed into multiple VNFCs). This option takes full advantage of the NFV potential at the expense of VNF-latency increment.
- Data Plane: Involves SDN switches used for packet forwarding. Data plane functionalities of S-GW and P-GW have been substituted by a Regional Router (RR) and regional switch (both SDN switches) that act as mobility anchors.

IV. DESIGN PROPOSALS

In this section, proposed handover and automated NS scaling procedures are given. Additionally, a virtualized EPC (vEPC) descriptor is realized to apply it the scaling procedures.

A. SDN-based X2 Handover

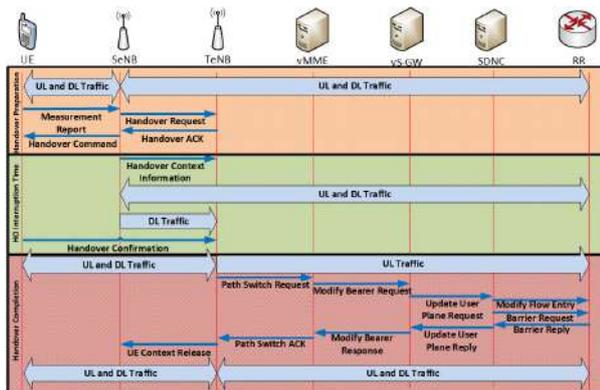


Figure 5: X2 Handover based on SDN

The proposed SDN handover procedure is carried out when a user moves between coverage areas of eNBs. This procedure is based on X2 handover of LTE [19], assuming that Mobility Management Entity (MME) and S-GW are not relocated. Fig. 5 depicts the workflow between EPC entities. This procedure is descomposed into three phases:

- Handover Preparation: If the signal level from target eNB overcomes an threshold, source eNB ask target eNB

if there are enough radio resources. If there are them, source eNB decides perform a handover.

- Handover Interruption Time: UE carries out a synchronization process with target eNB. During this period the UE cannot send or receive any packet. To support lossless HO, source eNB provide the sequence numbers of the forwarded packets.
- Handover Completion: Unlike previous phases, it differs regarding LTE handover. When a patch switch is requested, the SDN controller must take part to change the forwarding paths in the data plane. To that end, the vS-GW sends SDN controller a message to modify the corresponding flow table entry of SDN switches.

The SDN and NFV integration in this procedure impacts in the way LTE handover was performed. These impacts are summarized as follow:

- Data Plane: MPLS tunnels are used instead of GTP tunnels. MPLS is supported by OpenFlow protocol, and it reduces considerably the overhead size. Furthermore, each eNB and the RR have an associated SDN switch, called Edge Network Element (ENE), which push a MPLS labels in packets to indicate them the target ENE. When packets reach the target ENE, it pop their MPLS labels, and delivers them to target eNB.
- X2 Interface: The eNBs are interconnected with each other through X2 interface [19]. Each eNB stores a table that relates the Physical Cell Identifier (PCI), Evolved Cell Global Identifier (ECGI) and the IP address of its neighbouring eNBs. An eNB will learn the IP addresses of its neighbouring eNBs by requesting them directly to the SDN Controller, sending it the PCI.

B. NFV-based Automated Network Service Scaling

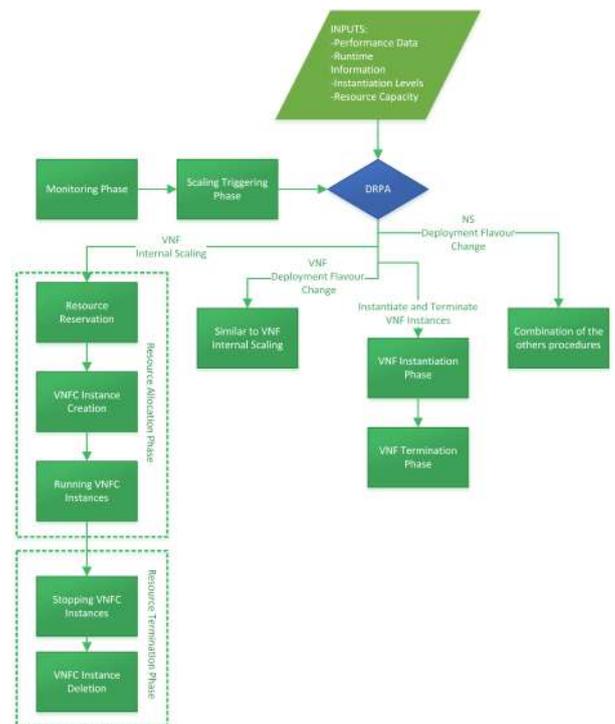


Figure 6: Automated Network Service Scaling Overview

The main goal of automated NS scaling procedures are the dynamic adaptation of NS capacity regarding NS incoming load. There are different procedures depending on abstraction level to be scaled:

- VNF Internal Scaling: Each VNF instance within a NS instance might be scaled. It involves adding/removing VNFC instances.
- VNF Deployment Flavour Change: If a VNF instance has been already scaled to its limits, the capacity increment/decrement is carried out through a deployment flavour change. Creation/deletion/modification of required external VLs is required along with the addition/deletion of VNFC instances.
- Instantiate and Terminate VNF instances: When there is no possibility of changing the deployment flavour of a VNF instance, VNF instances are instantiated or terminated.
- NS Deployment Flavour Change: If it is not possible adding or deleting VNF instances, the deployment flavour of the NS instance will be changed to a new one.

To perform one of these procedures, a set of scaling phases must be executed (see Fig. 6):

- Monitoring Phase: NFVO collects NS performance data.
- Scaling Triggering Phase: NFVO will use the collected performance data along with runtime information (current state of NS), instantiation levels (possible ways of NS deployment) and available resource capacity to decide how to scale a NS. This decision is taken using a Dynamic Resource Provisioning Algorithm (DRPA), which optimizes the scaling of a NS instance according to a set of criteria given in the correspond NSD. This algorithm outputs and optimum instantiation level and the selected VIM(s) for the resource management. Depending on the differences between optimum instantiation level and the previous one, one of the scaling procedures mentioned before will be applied.

Depending on triggered procedure, different steps are required. They are summarized below.

1) *VNF Internal Scaling*: To offer a seamless NS, resource allocation phase is performed before resource termination phase. In the former phase, a resource reservation is done in correspond VIM(s). After, new VFNC instances are created, and finally they are run. In the last phase, unnecessary VNFC are terminated.

2) *VNF Deployment Flavour Change*: This procedure is practically the same as VNF Internal Scaling. The only difference lies in the addition of new external VLs.

3) *Instantiate and Terminate VNF Instances*: VNF instantiation is performed before VNF instantiation to offer a seamless NS. VNF Instantiation phase is similar to resource allocation phase, with the singularity that all VNF's VNFCs must be created for the first time. VNF Termination is also similar to resource termination phase, with the singularity that all VNF's VNFCs must be terminated.

4) *NS Deployment Flavour Change*: This procedure basically combines the previous procedures. This combination will depend on how NSD is specified.

Each phase contains a sequence of messages<sup>1</sup> between NFV-MANO blocks [20] [21] [22] [23].

### C. Network Service Scaling Use Case: Virtualized EPC Scaling

To previously deploy a NS which acts as a vEPC, a NSD which describes the EPC is needed. In this work, a detailed description of a vEPC has been provided in the form of NSD<sup>1</sup>. Due to the great variety of EPC entities which constitute a vEPC, only a virtualized MME (vMME) description has been provided as VNFD. Additionally, the scaling procedures explained before have been applied to this vEPC<sup>1</sup>.

## V. EVALUATION

In this section, the handover procedure and the DRPA included in automated NS scaling procedures are evaluated.

### A. Handover Experimental Setup

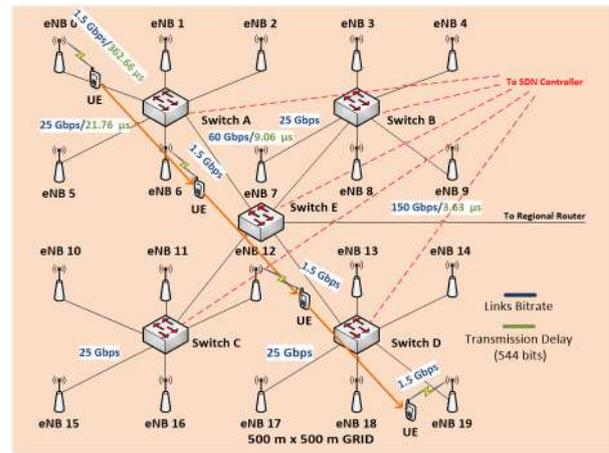


Figure 7: Handover Experimental Setup Topology

A simulator of the SDN handover procedure has been developed using ns-3 environment [24]. It implements the data plane, and the messages exchange and processing of network entities. Using this simulator, the handover procedure has been verified and its latency has been measured. The parameters given in Table I and the scenario depicted in Fig. 7 has been used.

### B. Handover Evaluation

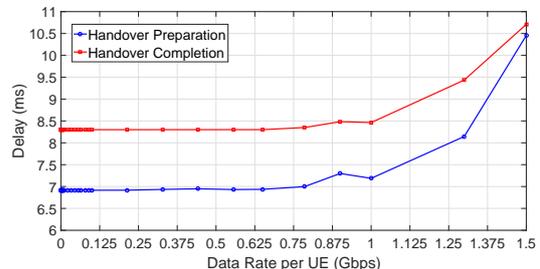


Figure 8: Handover Execution Time

Considering previous experimental setup, the handover execution time has been calculated. This latency is the sum of delays shown in Fig. 8 and the handover interruption time (constant).

According to [25], the control plane delay budget for LTE handover preparation and completion phases is 31 ms. Also,

Table I: Handover Experimental Setup Configuration

Network topology	
eNBs layout	Regular Grid 500 m x 500 m
eNB coverage area	100 m x 125 m
Number of eNBs	20
Position of the eNB $i \in \{0, \dots, 19\}$	$[50 + (i\%5) \cdot 100, 62.5 + [i/5] \cdot 125]$
Number of UEs	100
Backhaul topology	Tree (2 Levels)
Position of the BN switches	Square vertices and center (250,250). Side length: 250 m
RR position	25 Km from Switch E
UE Mobility	
Mobility model	Fluid-flow model
UE speed	6 m/s
Traffic model (UP)	
Traffic model type	ON-OFF model
ON and OFF periods	Uniformly distributed (0, 1) seconds
Processing delays	
eNB	960 $\mu$ s
OF Switch without modifying headers	5 $\mu$ s
OF Switch modifying headers	10 $\mu$ s
SDN Controller	3 ms
vSGW	1 ms
vMME	1 ms
Regional Router	100 $\mu$ s
Propagation delays	
Speed of light in wireless and wired links	300000 km/s
Handover measurement reports	
Event type	A3 with <i>offset</i> = 2 dB and <i>hysteresis</i> = 0dB
Carrier frequency	2.12 GHz
eNB Tx Power	30 dBm
Path loss model	Friis
Handover cases (probability)	
Intra-Switch Handover	80/126
Inter-Switch Handover	46/126

it is expected that 5G latency requirements will be two times more rigid in this plane [26]. Based on the above, the delay obtained meets the 5G latency requirements.

### C. MME Scaling Experimental Setup

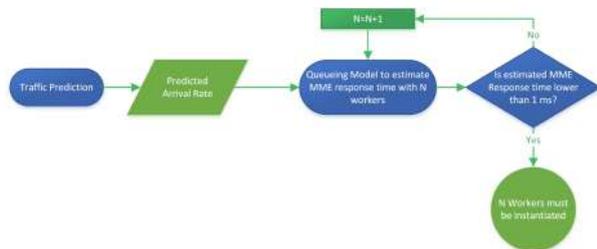


Figure 9: Proposed Dynamic Resource Provisioning Algorithm (DRPA)

To performance the vMME scaling, NFVO must execute the DRPA shown in Fig. 9. Firstly, incoming control plane traffic must be predicted. A perfect prediction of traffic evolution given in [27] is supposed. Then, following a MME queue model [28] given in Table II, the vMME response time is estimated to  $N$  workers (a VNFC type). This estimation is iteratively calculated for different values of  $N$  until it is less than 1 ms (required number of workers).

Table II: MME Scaling Experimental Setup Configuration

Queuing Model	
<b>Procedure</b>	<b>Mean Arrival Rate</b>
Service Request	$\lambda_{SR} = 0.0045$
S1 Release	$\lambda_{S1R} = 0.0045$
SDN-based Handover	$\lambda_{X2H} = 0.034375$
Aggregate	$\lambda = \lambda_{SR} + \lambda_{S1R} + \lambda_{X2H}$
<b>Model</b>	<b>Function</b>
Traffic Load Evolution	$p(t)$ See [27]
Mean Arrival Rate	$\lambda_{daily}(t) = 3 \cdot 10^5 \cdot \lambda \cdot p(t)$
<b>Procedure</b>	<b>Number of CPU Instructions</b>
Service Request	$I_{SR} = 2520000$
S1 Release	$I_{S1R} = 3200000$
SDN-based Handover	$I_{X2H} = 2140000$
<b>VNFCs</b>	<b>Mean Service Rate</b>
Load Balancer	$\mu_{LB} = 120000$
Workers	$\mu_W = \frac{1.96 \cdot 10^9 \cdot \lambda}{\lambda_{SR} \cdot I_{SR} + \lambda_{S1R} \cdot I_{S1R} + \lambda_{X2H} \cdot I_{X2H}}$
Database	$\mu_{DB} = 100000$
<b>VNFCs</b>	<b>Mean Response Time</b>
Load Balancer	$\bar{T}_{LB} = \frac{(\mu_{LB})^{-1}}{1 - \lambda_{daily}(t)/\mu_{LB}}$
Workers	$\bar{T}_W = (\mu_W)^{-1} + \frac{C(m, \rho)}{m \cdot \mu_W - \lambda_{daily}(t)}$ See Note
Database	$\bar{T}_{DB} = \frac{(\mu_{DB})^{-1}}{1 - \lambda_{daily}(t)/\mu_{DB}}$
Entire MME	$\bar{T}_{MME} = \bar{T}_{LB} + \bar{T}_W + \bar{T}_{DB}$
<b>Note:</b> <b>Erlang C Equation</b>	$C(m, \rho) = \frac{\left(\frac{m \cdot \rho}{1 - \rho}\right)^m}{\sum_{k=0}^{m-1} \frac{(m \cdot \rho)^k}{k!} + \left(\frac{m \cdot \rho}{1 - \rho}\right)^m} \cdot \left(\frac{1}{1 - \rho}\right)$

### D. Virtualized MME Scaling Evaluation

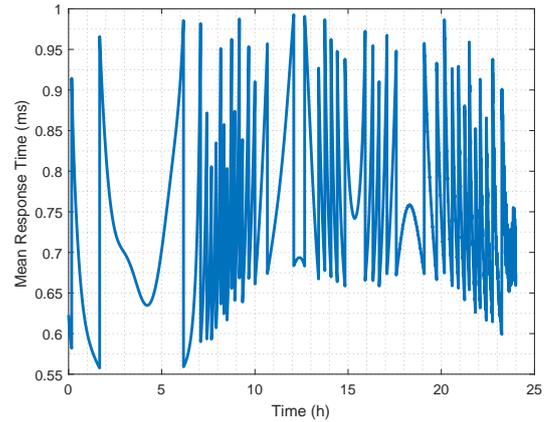


Figure 10: Mean Response Time of vMME

The vMME mean response time is shown in Fig. 10. This time never exceed 1 ms, thus the DRPA works properly. Furthermore, when traffic load slope increase (see [27]), there are an increment in the mean response time variability, whereas the opposite happens when the traffic load slope decrease.

## VI. CONCLUSION

In this work, an SDN handover procedure and automated NS scaling procedures have been proposed. On one hand, it has been verified that SDN handover works properly and its latency fulfill the 5G latency requirements. On the other hand, this work shed light on the NS scaling procedures. To automatically trigger any of them, the NFVO runs a DRPA which selects the optimum instantiation level as ETSI information model imposes. This fact bounds the scaling options, subjected to the discrete set of instantiation levels defined in the NSD. Additionally, to elucidate these procedures, a

NSD for the vEPC has been proposed, used it to illustrate the scaling a vEPC instance, specifically a vMME.

#### ACKNOWLEDGMENT

I would like to thank Pablo Ameigeiras Gutiérrez and Juan José Ramos Muñoz for their guidance. Furthermore, although he has not been formally my supervisor, I am deeply indebted to Jonathan Prados Garzón for their valuable comments on this work.

#### REFERENCES

- [1] V. G. Nguyen, A. Brunstrom, K. J. Grinnemo, and J. Taheri, "SDN/NFV-Based Mobile Packet Core Network Architectures: A Survey," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1567–1602, thirdquarter 2017.
- [2] E. Hernandez-Valencia, S. Izzo, and B. Polonsky, "How will NFV/SDN transform service provider OPEX?" *IEEE Network*, vol. 29, no. 3, pp. 60–67, May 2015.
- [3] J. Prados-Garzon, O. Adamuz-Hinojosa, P. Ameigeiras, J. J. Ramos-Munoz, P. Andres-Maldonado, and J. M. Lopez-Soler, "Handover implementation in a 5G SDN-based mobile network architecture," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2016, pp. 1–6.
- [4] O. Adamuz-Hinojosa, "SDN and NFV in 5G Networks: Handover and Automated Network Service Scaling as Uses Cases," Master's thesis, University of Granada, 2017.
- [5] D. Kreutz, F. M. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14–76, 2015.
- [6] O. S. Consortium *et al.*, "Openflow Switch Specification Version 1.0.0," 2009.
- [7] G. N. Operators, "Network Functions Virtualization: An Introduction, Benefits, Enablers, Challenges, & Call for Action," *white paper*, Oct, 2012.
- [8] *ETSI GS NFV-SWA 001 V1.1.1 Network Functions Virtualisation (NFV); Virtual Network Functions Architecture*.
- [9] D. ETSI, "NFV-EVE 005. Network Function Virtualization (NFV); Ecosystem; Report on SDN Usage in NFV Architectural Framework," 2015.
- [10] R. Mijumbi, J. Serrat, J. I. Gorricho, S. Latre, M. Charalambides, and D. Lopez, "Management and orchestration challenges in network functions virtualization," *IEEE Communications Magazine*, vol. 54, no. 1, pp. 98–105, January 2016.
- [11] N. ETSI, "GS NFV-MAN 001 V1.1.1 Network Function Virtualisation (NFV); Management and Orchestration," 2014, [http://www.etsi.org/deliver/etsi\\_gs/NFV-MAN/001\\_099/001/01.01\\_01\\_60/gs\\_nfv-man001v010101p.pdf](http://www.etsi.org/deliver/etsi_gs/NFV-MAN/001_099/001/01.01_01_60/gs_nfv-man001v010101p.pdf).
- [12] *ETSI GS NFV-IFA 014 V2.1.1 Network Functions Virtualisation (NFV); Management and Orchestration; Network Service Templates Specification*.
- [13] *ETSI GS NFV-IFA 011 V2.1.1 Network Functions Virtualisation (NFV); Management and Orchestration; VNF Packaging Specification*.
- [14] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 236–262, 2015.
- [15] M. Liyanage, A. Gurtov, and M. Ylianttila, *Software Defined Mobile Networks (SDMN): Beyond LTE Network Architecture*. John Wiley & Sons, 2015.
- [16] T. Taleb, M. Corici, C. Parada, A. Jamakovic, S. Ruffino, G. Karagiannis, and T. Magedanz, "EASE: EPC as a service to ease mobile core network deployment over cloud," *IEEE Network*, vol. 29, no. 2, pp. 78–88, 2015.
- [17] J. Costa-Requena, J. L. Santos, V. F. Guasch, K. Ahokas, G. Premsankar, S. Luukkainen, O. L. Pérez, M. U. Itzazelaia, I. Ahmad, M. Liyanage *et al.*, "SDN and NFV integration in generalized mobile network architecture," in *Networks and Communications (EuCNC), 2015 European Conference on*. IEEE, 2015, pp. 154–158.
- [18] P. Ameigeiras, J. J. Ramos-Munoz, L. Schumacher, J. Prados-Garzon, J. Navarro-Ortiz, and J. M. Lopez-Soler, "Link-level access cloud architecture design based on SDN for 5G networks," *IEEE Network*, vol. 29, no. 2, pp. 24–31, 2015.
- [19] T. ETSI, "ETSI TS 123 401 V8.14.0. LTE; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (3GPP TS 23.401 version 8.14.0 Release 8)," 2011, [http://www.etsi.org/deliver/etsi\\_ts/123400\\_123499/123401/08.14.00\\_60/ts\\_123401v081400p.pdf](http://www.etsi.org/deliver/etsi_ts/123400_123499/123401/08.14.00_60/ts_123401v081400p.pdf).
- [20] *ETSI GS NFV-IFA 005 V2.1.1 Network Functions Virtualisation (NFV); Management and Orchestration; Or-Vi reference point - Interface and Information Model Specification*.
- [21] *ETSI GS NFV-IFA 006 V2.1.1 Network Functions Virtualisation (NFV); Management and Orchestration; Vi-Vnfm reference point - Interface and Information Model Specification*.
- [22] *ETSI GS NFV-IFA 007 V2.1.1 Network Functions Virtualisation (NFV); Management and Orchestration; Or-Vnfm reference point - Interface and Information Model Specification*.
- [23] *ETSI GS NFV-IFA 008 V2.1.1 Network Functions Virtualisation (NFV); Management and Orchestration; Ve-Vnfm reference point - Interface and Information Model Specification*.
- [24] G. F. Riley and T. R. Henderson, "The ns-3 network simulator," in *Modeling and Tools for Network Simulation*. Springer, 2010, pp. 15–34.
- [25] Z. Li and M. Wilson, "User plane and control plane separation framework for home base stations," *Fujitsu Scientific and Technical Journal*, vol. 46, no. 1, pp. 79–86, 2010.
- [26] F. Hu, *Opportunities in 5G networks: A research and development perspective*. CRC Press, 2016.
- [27] T. Louail, M. Lenormand, O. G. Cantú, M. Picornell, R. Herranz, E. Frias-Martinez, J. J. Ramasco, and M. Barthelemy, "From mobile phone data to the spatial structure of cities," *arXiv preprint arXiv:1401.4540*, 2014.
- [28] J. Prados, J. Ramos-Munoz, P. Ameigeiras, P. Andres-Maldonado, and J. Lopez-Soler, "Modeling and Dimensioning of a Virtualized MME for 5G Mobile Networks," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2016.



**Óscar Adamuz Hinojosa** received his B.Sc. and M.Sc. in Telecommunications engineering in 2015 and 2017, respectively, from the University of Granada, Spain. He is currently working toward the Ph.D. degree with the Department of Signal Theory, Telematics and Communications, University of Granada, Spain. His research interest include Software Defined Networking, Network Functions Virtualization, and Network Slicing in 5G.

# Urban 5GRX: Herramienta para el despliegue de 5G en entornos urbanos densos

Autor: Francisco Jesús Quero de la Rosa, e-mail: franciscoquero@gmail.com

Tutores: Pablo Muñoz Luengo, e-mail: pabloml@ugr.es, Juan Francisco Valenzuela Valdés, e-mail: juanvalenzuela@ugr.es

Titulación: Máster Universitario en Ingeniería de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—En este proyecto se ha construido una herramienta de despliegue para redes heterogéneas 5G en entornos densos, con información detallada y real de cualquier punto geográfico acerca de las redes ya existentes en ese punto, así como las condiciones del entorno, como por ejemplo edificios y elevación del terreno. Para su desarrollo, el proyecto se ha apoyado en el uso de una novedosa *toolbox* de Matlab que fue lanzada recientemente y que permite la visualización de terrenos a partir de Open Street Map. Además, se llevan a cabo varias simulaciones de despliegue en la localidad de Granada, ciudad donde no existe despliegue de 5G a la fecha de realización del proyecto, con la finalidad de evaluar los resultados de un hipotético despliegue.

**Palabras clave**—5G, Alta frecuencia, Comunicaciones de Radio, Comunicaciones Inalámbricas, Despliegue, GIS, HetNets, Macro-Celdas, Matlab, Micro-Celdas, Planificador, Redes 5G, RF Toolbox, Social attractors.

## I. INTRODUCCIÓN

**D**ESPLEGAR infraestructuras de tecnologías de nuevas generaciones puede resultar una tarea que requiere cierta cantidad de recursos con importantes costes asociados, tanto de infraestructura como de mano de obra [2]. Además, debido a la creciente complejidad de las mismas, el desarrollo requiere una mayor atención, especialmente en escenarios heterogéneos (*Heterogeneous Networks, HetNets*), Figura 1, como lo es el estándar 5G, cuyo punto fundamental es justamente la coexistencia de distintos tipos de estaciones base (*Base Stations, BSs*).

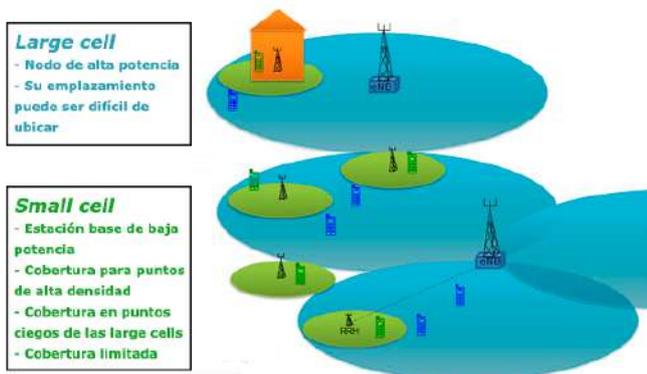


Fig. 1. Entorno típico de HetNet [3]

En un escenario de redes heterogéneas, podemos encontrar dos tipos de celda: macro-celdas y micro-celdas, que a su vez

se pueden clasificar con otros criterios en diversos subtipos. Cada tipo de celda cuenta con una finalidad distinta y un ecosistema de todas ellas se puede beneficiar no solo de una cobertura más extensa, sino también de un abastecimiento suficiente para el uso intensivo de las redes móviles que se espera según las tendencias, y es que se espera que el tráfico móvil se triplique antes de que el año 2025 finalice, cuando, además, la red 5G soportará el 50% de todo este tráfico [2].

En la actualidad, existe mucho trabajo por hacer con respecto al despliegue de 5G. A fecha de junio de 2020, tan solo 22 despliegues de redes 5G se han sido realizados en España.



Fig. 2. Mapa de puntos de acceso 5G en la Península Ibérica a fecha de junio de 2020 [1]

La tarea de despliegue viene condicionada por la demanda de los usuarios, puesto que se requiere cubrir la demanda de utilización que se ha comentado anteriormente. Estas labores se complican incluso más cuando no existen directrices estándares o proporcionadas por ninguna entidad con autoridad, por lo que es responsabilidad de cada operador realizar un despliegue lo más optimizado posible, sin desperdiciar recursos como estaciones base colocadas mientras que se consigue proporcionar un correcto servicio a los usuarios, y suficiente para su demanda de uso.

Por ello, las herramientas de despliegue toman más pro-

tagonismo cuando la complejidad de las tecnologías aumenta. Dicho esto, con este Trabajo de Fin de Máster se pretende elaborar una herramienta de despliegue de redes 5G para el caso de uso concreto de los entornos urbanos, con el propósito de que pueda servir para conocer los requisitos que una futura infraestructura deben cumplir para un correcto funcionamiento.

## II. RF TOOLBOX DE MATLAB

Con fecha de 19 de marzo de 2019, Matlab lanza su versión R2019a y, con ella, despliega una serie de nuevas funcionalidades. Entre ellas, se encuentra la llamada *RF toolbox*, la cual ofrece funciones, objetos y aplicaciones para diseñar, modelar, analizar y visualizar redes y componentes de radiofrecuencia. En el lanzamiento de la versión R2019a, Matlab incluye una función muy llamativa, la cual permite visualizar en el mapa los elementos de las redes que el usuario genere, mostrándolos a través de una interfaz que muestra el terreno.

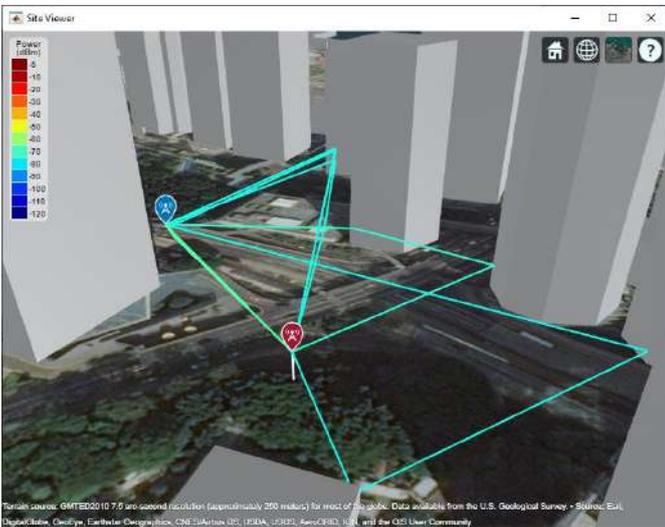


Fig. 3. Ejemplo de visualización de las diferentes líneas de trazado de los rayos con sus reflexiones entre los edificios.

Otra de las funcionalidades que añade potencial es la de añadir edificios al mapa mostrado por la herramienta, lo que permite que se consideren a la hora de realizar los cálculos. Sin embargo, no existe una forma nativa de añadir los edificios, puesto que deben de cargarse desde un fichero previamente almacenado en el equipo.

La abstracción que ofrece esta *toolbox* ofrece la suficiente potencia como para ser utilizada como base para herramientas más completas. En concreto, se buscará combinar el potencial de software geográfico con el del cálculo de las propagaciones, combinando así las funcionalidades para simulaciones precisas de la trayectoria de las señales.

## III. DISEÑO DE URBAN 5GRX

Urban 5GRX surge a partir de todo lo que se ha comentado en capítulos anteriores. Su intención es la de cubrir la necesidad de resolver el rompecabezas del despliegue de una tecnología tan compleja como 5G en el escenario más intrincado como lo es el del urbano denso.

Puesto que la tecnología 5G no se encuentra desplegada actualmente, no solo no es posible encontrar información pública de los actuales emplazamientos, ya que no existen, sino que los encargados de los despliegues se encuentran con una hoja en blanco en la que deben elegir bien dónde situar los nuevos puntos de acceso. Por tanto, haciendo referencia a los distintos modos de despliegue inicial que Ericsson [2] propone, el que menor inversión inicial supone y el que más sentido tiene a la hora de iniciar un despliegue completo es el modo 2. Este modo consiste en emplazar los nodos de 5G en nodos LTE ya existentes, a la vez que se emplazan micro celdas en puntos estratégicos para garantizar al usuario una buena relación cobertura-capacidad.

El propósito de esta herramienta es la de modelar las comunicaciones de una forma lo más realista posible, incluyendo características del entorno como edificios o altitud, modelado de distintos tipos de celdas que componen la red heterogénea que dará servicio, puntos de concentración de usuarios, un emplazamiento óptimo para los distintos puntos de acceso y, en general, todo parámetro que pueda suponer la diferencia entre un peor o un mejor rendimiento.

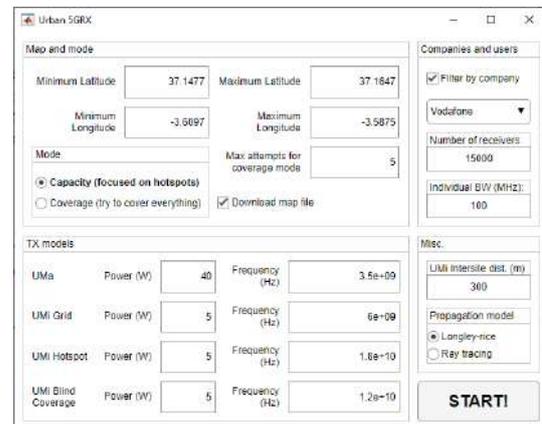


Fig. 4. Interfaz gráfica de Urban 5GRX antes de ejecutar una simulación.

Para facilitar esta tarea, se ha concebido esta herramienta como una aplicación automática, con interfaz gráfica para que el usuario no necesite aprender a utilizarla a través de código. Configurar parámetros y hacer clic en el botón de comienzo basta para realizar la simulación del despliegue en una zona determinada.

Además, se incluye una funcionalidad de exportación de resultados a un fichero de texto para que pueda consultarse posteriormente así como dejar abierta la posibilidad de analizarlo gramaticalmente a través de alguna herramienta automática de modo que pueda interpretar los resultados y tomar acciones a partir de ellos, como importarlos a un simulador u otra herramienta de despliegue.

### A. Características

- Cuenta con una intuitiva interfaz gráfica que simplifica el uso de la herramienta hasta el punto en el que solo hace falta clicar en un botón para lanzar la ejecución.
- Descarga automáticamente los modelos 3D de los edificios de la zona seleccionada. Los tiene en cuenta para simular la propagación de la señal.

- Descarga automáticamente los emplazamientos de los puntos de acceso ya existentes para LTE. Permite seleccionar una operadora entre las tres principales en España (Vodafone, Movistar y Orange) para filtrar para que solo se muestren las que pertenezcan a la compañía seleccionada. Estos emplazamientos se utilizan para situar los nodos de 5G aprovechando la infraestructura y partiendo de la premisa de que, en muchos casos, es posible actualizar los equipos existentes para compatibilizarlos con 5G. Sin embargo, no utiliza todos los emplazamientos sino aquellos alejados entre sí más de 100 metros de distancia.
- Crea una red heterogénea con hasta cuatro tipos de celda distintos: UMa, UMi de alta capacidad (hotspot), UMi de cobertura (muchos puntos de acceso que se distribuyen uniformemente por la superficie) y UMi de cobertura de puntos ciegos. El modelo distingue tipo de antena, altura, potencia y frecuencia. Los emplazamientos para todos los nodos se generan automáticamente.
- Dos modos de funcionamiento: uno enfocado a cubrir el 100 % de la superficie y el segundo enfocado a poner énfasis en cubrir los puntos que mayor concentración de personas tienen.
- Identifica automáticamente los tipos de edificio que existen en el área seleccionada y los clasifica, identificando los puntos de mayor interés para emplazar allí una mayor concentración de usuarios.
- Permite la simulación de hasta 15.000 usuarios en el mismo mapa.
- Cuenta con dos tipos distintos de propagación: *Longley Rice* y *Ray Tracing Image Method*. *Longley Rice* es más rápido pero solo tiene en cuenta factores como si existe LOS o no, la difracción y la dispersión, todo debido a los obstáculos. *Ray tracing* es un modelo más detallado y, por tanto, más lento. Realiza un estudio detallado sobre la trayectoria de cada rayo de señal, con un número determinado de reflexiones, para determinar el camino que seguiría la señal y las posibles atenuaciones, como absorción por rebote o visión directa.
- Permite frecuencias de hasta 100 GHz.
- Determina enlaces de backhaul entre micro celdas y macro celdas, asumiendo que las micro-celdas dependerán de las macro celdas para el abastecimiento estando conectadas entre sí, mientras que se asume que las macro celdas están conectadas a la red troncal cableada.
- Determina la capacidad y la SINR para cada usuario, con la finalidad de conocer de cuánto ancho de banda se ha de abastecer en cada zona.
- Calcula el porcentaje de superficie donde la señal recibida tendría una SINR mayor que cero, es decir, calcula el porcentaje de cobertura para toda la superficie.
- Realiza una exportación de resultados para que puedan ser analizados en detalle después de su ejecución.
- Cuenta con una visualización del mapa y de la red mejorada, con la que se puede distinguir, a simple vista y con el color, el tipo de celda de cada emplazamiento, la distribución de usuarios, el nivel de señal de cada punto del mapa, los enlaces de backhaul y más aspectos.
- Se incluye un *script* independiente que permite la opti-

mización de los ángulos de los distintos sectores de las UMa, puesto que están compuestas por tres celdas sectoriales, cada una a un ángulo distinto. Dicho *script* utiliza algoritmos genéticos para encontrar la combinación de celdas que otorguen la mejor cobertura de la zona.

#### B. Limitaciones

- Las frecuencias de trabajo máximas del modelo de propagación *Longley-rice* son 20 GHz, lo que puede ser insuficiente a la hora de simular ciertos tipos de UMi. Sin embargo, la frecuencia máxima para el modelo de *Ray tracing* es de 100 GHz.
- Los tiempos de ejecución para el modelo de *Ray tracing* se pueden prolongar hasta varios días en un caso con numerosas estaciones base y una superficie grande.
- No se pueden simular escenarios con más de 15.000 usuarios.
- Debido a limitaciones de la fuente de la que se descarga los mapas, la máxima superficie que se puede estudiar en una sola simulación es de 4 km<sup>2</sup>.
- En Urban 5GRX no existe una forma de encontrar las coordenadas de la superficie a calcular, por lo que se ha de elegir una herramienta externa, como Open Street Map, para seleccionar un área y extraer las coordenadas deseadas.
- No es posible modificar el modelo de las antenas sin modificar el código.
- Esta herramienta asume que el ancho de banda asignado a cada usuario se consigue siempre. No tiene en cuenta técnicas de diversidad, modulaciones u otros detalles de bajo nivel que se tendrían que aplicar para garantizar dicho ancho de canal.
- Como herramienta automática, no es posible seleccionar manualmente los emplazamientos de las estaciones base y/o usuarios.

#### IV. IMPLEMENTACIÓN DE URBAN 5GRX

La herramienta Urban 5GRX ha sido implementada dividiendo sus funcionalidades en módulos que han sido desarrollados por separado, para así poder mejorar la escalabilidad y simplificar el mantenimiento a largo plazo de la herramienta. En concreto, está construida con los siguientes componentes:

- **Conector de Open Street Map.** Este conector es la base sobre la que se alza Urban 5GRX. Se encarga de descargar ficheros de modelos de edificios en 3D que pueden ser representados en el visor nativo de Matlab. Puesto que este planificador se basa en el supuesto de ofrecer al usuario la posibilidad de utilizar los datos reales del entorno de cualquier punto del globo terráqueo, y para ello, existe un límite y es que Matlab no es capaz de extraer esos edificios por sí solo. Por ello, este conector está dedicado a obtener ese fichero de la red.
- **Conector de OpenCellID.** El objetivo de este conector es el de obtener la localización de los puntos de acceso ya existentes para LTE. Las coordenadas de los mismos podrán usarse para reutilizar la localización para desplegar macro-celdas.

- **Generador de Macro-Celdas.** Partiendo de las posiciones que se han calculado para situar las macro-celdas, el siguiente paso es el de generarlas, utilizando este componente que no solo las genera de forma que Matlab puede procesarlas, sino que modela de acuerdo a parámetros ofrecidos por el 3GPP para el uso experimental de redes 5G. En la Fig. 5 se puede observar el modelado de la antena diseñada para el uso en macro-celdas.

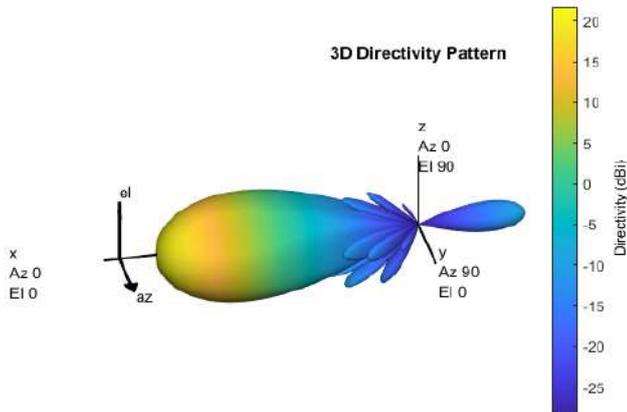


Fig. 5. Patrón de radiación de la antena utilizada en cada sector de macro-celdas para 3,5 GHz.

- **Evaluador del entorno.** Este módulo realiza mediciones de forma distribuida por todo el mapa dentro de los límites establecidos por el usuario. Estas mediciones en forma de malla toman el valor del mejor valor de la SINR de ese punto (elegido entre todos los transmisores) y la devuelve junto a la latitud y longitud del punto de medición. Además, se añade un mapa de calor para representar visualmente los valores recibidos.
- **Optimizador de la orientación de sectores.** A pesar de que por defecto, los sectores de cada macro-celda están separados entre sí  $120^\circ$ , esto es un problema potencial si todos tienen la misma orientación, ya que por directividad, dos estaciones base interferirán entre sí si están alineadas. Para abordar este problema, se ha realizado un *script* independiente que hace uso de algoritmos genéticos para determinar un conjunto de ángulos que garanticen un buen rendimiento de la red en cuanto a interferencias y cobertura se refiere.
- **Extractor de social attractors.** Como parte del módulo generador de usuarios, se ha diseñado un detector de puntos de interés escrito en Python llamado *Open Street Map Buildings Parser*. Estos puntos de interés se conocen en inglés como *social attractors*, y son puntos, lugares o edificios con una capacidad de concentrar personas más elevada que lo normal.
- **Generador de usuarios.** Este módulo genera receptores distribuidos en el mapa, repartidos por todo el mapa de forma no uniforme. Se crea un conjunto de N latitudes y longitudes, donde N es el número total de receptores introducido por el usuario. El método de generación utiliza las coordenadas de los edificios como referencia

de la generación de receptores, de modo que la distribución se realiza de forma heterogénea dependiendo de la importancia de cada edificio recibido del anterior componente.

- **Generador de Micro-Celdas.** Se ha concebido un total de tres tipos de micro-celdas para poder satisfacer los requisitos de ambos modos de ejecución. Cada tipo de celda tiene una serie de parámetros comunes, como su altura y el esquema de radiación de su antena, pero otros parámetros cambiarán según lo aportado por el usuario. Este módulo los genera y modela de acuerdo a los parámetros otorgados.
- **Generador de localizaciones de Micro-Celdas.** Lo que define en cierto modo cada tipo de micro-celda es su emplazamiento. Existen tres formas de generar localizaciones de micro-celdas en Urban 5GRX y es lo que diferencia una de otras por defecto. Este generador emplaza las micro-celdas generadas de acuerdo al modo seleccionado, de los que se dispone tres en total.
- **Visor mejorado.** Debido a la gran cantidad de elementos que se han de representar en el visor, se ha mejorado la forma en la que se representan para proporcionar más información a simple vista y hacer más fácil la interpretación de los elementos.

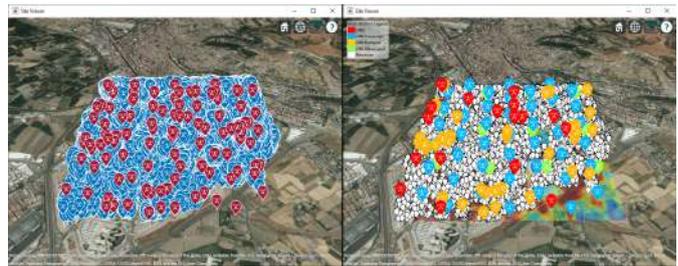


Fig. 6. Comparativa entre la visualización nativa (izquierda) y la visualización mejorada (derecha).

- **Calculadora de enlaces de backhaul.** De las decisiones del diseño de la red se decidió concebir las micro-celdas como sitios independientes cuyo acceso a la red troncal se hace por medio de un enlace inalámbrico hasta el lugar de una de las macro-celdas, ya que los únicos accesos a la red troncal por cable se producen desde estos emplazamientos. Para determinar hacia dónde debería estar orientado el enlace, se ha implementado un caso de uso que calcula la distancia entre cada micro-celda y cada macro-celda, y determina también si existe visión directa o no.
- **Exportación de resultados.** Como último paso de las ejecuciones, se ha implementado un módulo que exporta resultados e información importante de la planificación. Como paso previo, se realizan una serie de cálculos.
- **Interfaz gráfica de usuario.** Como última parte de la implementación de Urban 5GRX se ha querido prestar atención a la facilidad de uso. Para ello, se ha implementado una interfaz gráfica de usuario utilizando la herramienta de Matlab llamada Appdesigner, Figura 7.

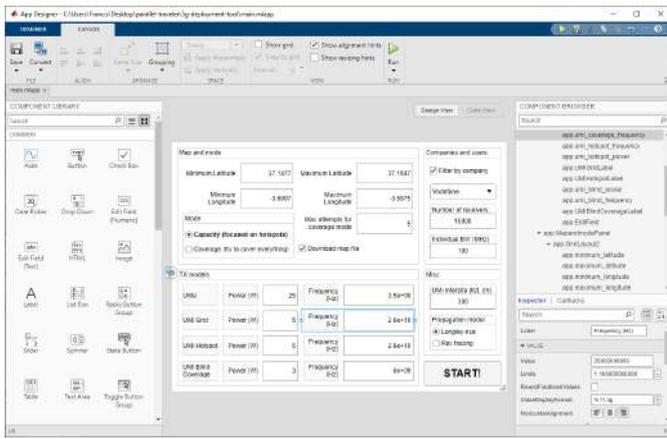


Fig. 7. Interfaz gráfica de Appdesigner.

## V. CASO DE ESTUDIO: DESPLIEGUE DE 5G EN GRANADA

Para finalizar con el proyecto, se ilustran los resultados de la planificación generada para dos escenarios reales. Ambos escenarios elegidos se encuentran en Granada, pero en barrios distintos. Por un lado, se desplegará en un barrio no céntrico y con mayor variedad de edificios, el Zaidín, y por otro lado, se realizará el estudio para la zona centro-Figueras.

### A. Evaluación para el barrio del Zaidín

La ubicación elegida para esta primera simulación es el barrio del Zaidín, situado al sur de la ciudad. El visor muestra el reparto poblacional por toda la zona seleccionada, que, como era de esperar, no es homogéneo, ya que existen zonas sin viviendas o edificios de especial concentración como un hospital, el estadio de fútbol Nuevo los Cármenes o el Parque Tecnológico de la Salud, sede universitaria.

Se puede hacer especial mención a las celdas naranjas, las de puntos de especial concentración de usuarios, que se encuentran precisamente en estos sitios más destacables, como cerca del estadio de fútbol o en el mismo tejado del hospital, aunque no se puedan apreciar exactamente en las imágenes por la gran concentración de usuarios.

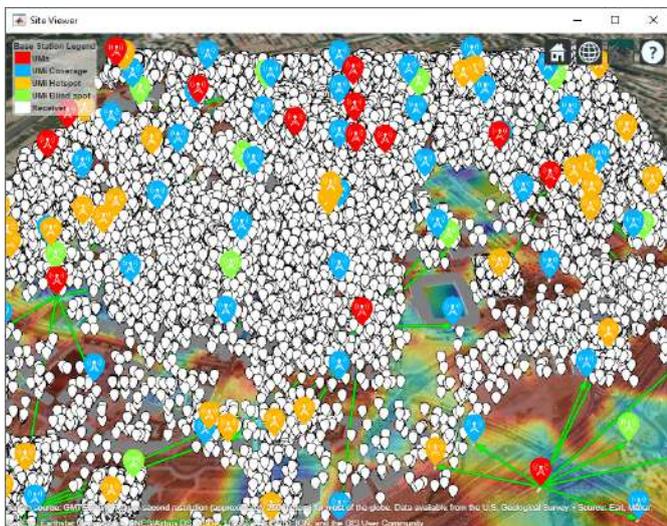


Fig. 8. Vista inclinada del escenario simulado.

Como datos destacables de los resultados de salida, mencionar que se está consiguiendo una cobertura total superior al 95 % y una capacidad media por usuario de unos 344 Mbps, lo cual cumple con los objetivos propuestos por todas las guías y recomendaciones de despliegue que se han consultado en anteriores capítulos.

### B. Evaluación para el centro de Granada

Del mismo modo que anteriormente, se repite una simulación. Esta vez, la zona seleccionada es el centro de Granada, concretamente el área alrededor de Acera del Darro, Carrera de la Virgen y el río. En la Fig. 9 se puede ver que, con vista aérea, la población se encuentra muy repartida por toda la zona, algo de esperar en lo que viene siendo el centro de Granada.



Fig. 9. Vista inclinada del escenario simulado.

De los datos obtenidos, se puede extraer que el resultado cumple con las expectativas con una cobertura mayor del 95 %, incluso mayor que en el anterior caso, con unos 331 Mbps por usuario y un tráfico total de 4,97 Tbps, algo más ligero que en el anterior caso.

Las diferencias de cifras entre ambos casos tiene explicación. Al haber mayor dispersión de usuarios, es más difícil dar cobertura, ya que no se encuentran tan concentrados en los mismos puntos. Por tanto, aunque en este caso haya una mejor cobertura en cuanto a zona cubierta, existen menos usuarios con una SINR alta, y el tráfico total del área decrece con respecto al anterior caso aunque haya el mismo número de usuarios, ya que hay mayor probabilidad de que se encuentren en puntos relativamente lejanos de sus puntos de acceso.

## VI. CONCLUSIONES

Durante el transcurso del proyecto, se ha desarrollado e implementado una herramienta de despliegue en entornos urbanos densos para 5G, llamada Urban 5GRX, totalmente funcional e implementada en Matlab y en Python 3. Se ha conseguido que esta herramienta sea intuitiva a la hora de su utilización gracias a su interfaz gráfica y al grado de automatización con el que cuenta. Además, gracias a sus funcionalidades, se pueden obtener representaciones gráficas

del entorno simulado ya que posee la capacidad de mostrar el entorno urbano simulado con representaciones en tres dimensiones de sus edificios, y con descripción visual de los diferentes elementos de la red y de su ubicación.

El planificador Urban 5GRX ha demostrado cumplir su propósito y ajustarse a los objetivos aun teniendo recursos y tiempo limitados para su consecución. La herramienta permite estudiar el hipotético despliegue en cualquier escenario del mundo con tan solo clicar en un botón, esperándose que este simulador pueda servir como base para un proyecto mucho más ambicioso que podría deparar en un producto completamente funcional.

#### AGRADECIMIENTOS

Agradecer a mis tutores Pablo y Juan Francisco la labor que durante el curso 2019/20 desempeñaron mentorizando el presente proyecto. Gracias a ellos no solo Urban 5GRX vio la luz, sino que han despertado en mí inquietudes en las que quiero seguir indagando.

#### REFERENCIAS

- [1] *OOKLA 5G Map™*, junio de 2020. Disponible: <https://www.speedtest.net/ookla-5g-map>
- [2] Ericsson. *5G deployment considerations*, 2020. Disponible: <https://www.ericsson.com/4a5daa/assets/local/networks/documents/5g-deployment-considerations.pdf>
- [3] Jeanette Wannstrom and Keith Mallinson. *Hetnet/small cells*, 2014.

# Aplicación móvil para detección y alerta de casos de *cyberbullying*

Autor: Beatriz Quesada Tobar, e-mail: beatrizquesada@correo.ugr.es

Tutor: Dr. D. Pedro García Teodoro, e-mail: pgteodor@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—La tecnología avanza a pasos agigantados, a un ritmo más rápido que las medidas de protección dirigidas a la población más vulnerable: niños y adolescentes. Si la tecnología permite facilitar las relaciones sociales entre los más jóvenes, también correspondería a esta dotar de las medidas necesarias para que su mal uso no repercuta en el desarrollo emocional de estos usuarios. En este marco, el presente proyecto persigue el desarrollo de una herramienta Android capaz de analizar si el usuario puede estar siendo víctima de acoso o presenta un estado de ánimo depresivo. En tal caso, los tutores legales podrán ser informados y tomar las medidas que estimen convenientes.

**Palabras clave**—Android, Ciberacoso, Python, Monitorización, Twitter.

## I. INTRODUCCIÓN

EN los últimos años, el teléfono móvil o *smartphone* se ha convertido en un objeto indispensable en la vida de toda persona. Lo que se diseñó en un comienzo como una herramienta para recibir y realizar llamadas, ha experimentado tal desarrollo que, actualmente, permite un sinnúmero de funcionalidades y resulta casi imposible completar una jornada sin él. No es de extrañar, por tanto, que las relaciones sociales y otros hábitos cotidianos se hayan trasladado ampliamente al sector de las TIC. Ante un desarrollo tecnológico de tales dimensiones es necesario que las normas y leyes que regulan su uso, así como la conciencia y responsabilidad de la población, avancen a la misma velocidad. Sin embargo, este no ha sido el caso. Y es que un uso indebido de estos dispositivos, y en especial de las redes sociales en población infanto-juvenil, pueden causar un deterioro en el bienestar de niños y adolescentes.

El ciberacoso, o acoso virtual, consiste en el uso de medios digitales, plataformas o redes sociales con el fin de “molestar” o dañar de manera intencionada a otra persona. El *cyberbullying* [1] ya representa uno de cada cuatro casos de acoso escolar y a partir de los 13 años el 36,5 %, es decir, uno de cada tres. Son las mujeres las que más lo sufren, 70 % frente al 30 % de los varones, realizándose en mayor medida a través del *smartphone*. En la Figura 1 puede observarse el amplio porcentaje que ocupa el ciberacoso frente a otros tipos de acoso y como casi la totalidad de estos abusos se cometen haciendo uso de los *smartphones*.

La adicción [3] a las redes sociales es otro de los grandes problemas de este siglo que merma, de manera significativa, la calidad de vida de pequeños y jóvenes. La adicción

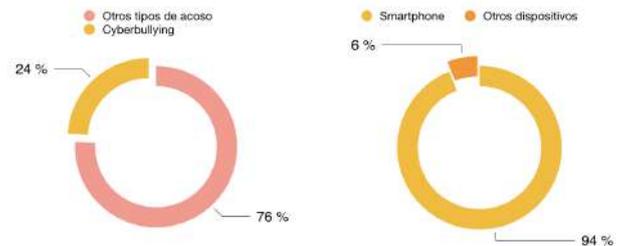


Fig. 1. Porcentaje de tipos de acoso y dispositivos en los que se realiza.

acaba por consolidarse cuando se produce un uso abusivo descontrolado, aparecen síntomas de abstinencia (ansiedad, depresión, intranquilidad) ante la pérdida temporal de conexión, se establece la tolerancia (es decir, la necesidad creciente de aumentar el tiempo de conexión) y de ahí derivan consecuencias negativas para la vida cotidiana.

En este proyecto se ha llevado a cabo el desarrollo e implementación de una herramienta TIC que permita a los padres o tutores legales constar de cierta información de la actividad de sus hijos en la red social Twitter. Está orientada a un control de los hábitos del menor que, una vez interpretados adecuadamente, podrán esclarecer si existe la posibilidad de que nos encontremos ante una situación, o no, que merezca una especial atención por parte de los tutores.

## II. SOLUCIONES EXISTENTES

Conforme se han ido presentando estos problemas, se han desarrollado numerosas herramientas o aplicaciones para intentar contrarrestar estas situaciones. Una de las soluciones que parecen más evidentes al hablar de estos problemas es dotar a los padres del control, total o parcial, del dispositivo de sus hijos/as, las llamadas herramientas de control parental. Proveer a los menores de un dispositivo como el *smartphone* no implica un mal uso del mismo, y puede utilizarse para remediar o afrontar “desde dentro” los distintos problemas. Esto es lo que pretenden las aplicaciones orientadas a los menores.

Las aplicaciones de control parental proporcionan a los tutores el control más absoluto del comportamiento de sus hijos. Expresado en otras palabras, sería como disponer del propio dispositivo. Esta solución resulta, claro está, muy pertinente. No obstante, ¿qué ocurre cuando el menor ya



Fig. 2. Logo de la herramienta VORDRA.

se encuentra en la etapa adolescente o pre-adolescente? Por todos es conocido que la adolescencia conlleva una necesidad intrínseca de una “parcela” propia de intimidad y privacidad, cuya vulneración desata sentimientos difíciles de manejar en esta edad. Justo es en este momento donde aumenta considerablemente la probabilidad de que aparezcan los problemas comentados previamente y, por tanto, sería el momento ideal de tener completo acceso a sus redes sociales. Sin embargo, las aplicaciones expuestas con anterioridad acarrear una de las mayores limitaciones: pueden ser percibidas por el menor como una violación de su intimidad, conllevando a una fácil desinstalación de las mismas. Ante esto, lo progenitores vuelven a quedarse a ciegas, sin hablar de la pérdida de la mucha o poca confianza que tenían con sus padres. Llegamos a la conclusión, por tanto, de que la herramienta desarrollada debe presentarse al menor como “inofensiva”, pudiendo asegurarse que no se tendrá acceso al contenido explícito de los mensajes intercambiados.

### III. DISEÑO DE LA HERRAMIENTA

El objetivo final ha sido la implementación de una herramienta que permita proporcionar a los tutores de un menor información sobre el uso de su hijo/a de una red social. El diseño de dicha herramienta deberá de ser completamente funcional y escalable.

La herramienta se ha denominado VORDRA haciendo referencia al significado en nórdico antiguo de VORDR: guardián, cuidador o protector. Esta herramienta cuenta con un logotipo propio mostrado en la Figura 2.

Se distinguen 4 puntos principales dentro de esta propuesta:

- Se propone la monitorización de la red social Twitter por ser una de las más utilizadas por jóvenes y adolescentes (incluida en el top 5 de redes sociales [2]). Además, la API de Twitter [4] ofrece múltiples posibilidades para adaptarla a diversos lenguajes y permite recopilar información del menor que sea útil. La herramienta, al contrario de las soluciones ya estudiadas, no es excesivamente intrusiva puesto que se presentará a los padres los resultados de un análisis elaborado a partir de la información de la cuenta del menor, sin mostrarles en contenido explícito. Se apuesta así por un acuerdo entre padres e hijos y que los menores no la visualicen como una amenaza a su intimidad.
- El cumplimiento de la ley de protección de datos debe estar garantizado en todo momento. Esto se hace con la obligación de la creación de una cuenta, de manera que, sólo introduciendo las credenciales de esa cuenta vinculada al menor, pueda accederse a la información.

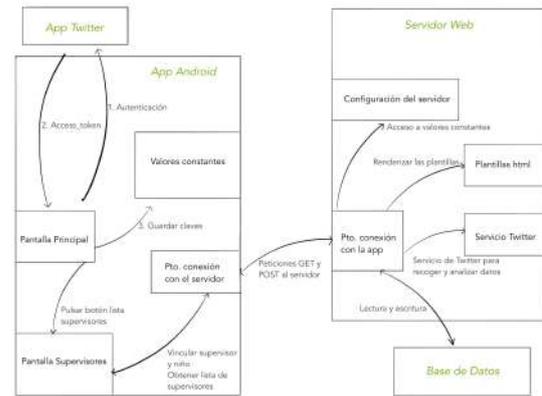


Fig. 3. Arquitectura general de VORDRA.

- Escalabilidad, debe que contemplarse la posibilidad de que, tanto un padre/madre tenga más de un hijo/a, como que un hijo/a tenga más de un supervisor/a.
- Se escoge una exposición visual de los datos. Es conveniente que los tutores puedan entender de manera fácil y rápida la información de la monitorización. En esta solución mediante gráficas, la representación de los resultados del análisis es muy intuitiva, permitiendo una visión general de la información.

### IV. IMPLEMENTACIÓN

La parte de desarrollo Android ha sido realizada con Android Studio versión 3.6.2 [5] en lenguaje Java. Además se ha hecho uso de la librería twitter4j [6] para llevar a cabo el proceso de autenticación en Twitter y la librería Retrofit [7] para la creación de un cliente de red, necesario para la comunicación con el servidor.

El desarrollo web se ha llevado a cabo con el entorno de programación Visual Studio Code [8] con ficheros en Python [9] y HTML principalmente. También ha sido necesario el uso de DB Browser SQLite [10] para almacenar los distintos parámetros.

El objetivo principal del desarrollo de una *app* en Android es la de permitir a dicha *app* el acceso a la cuenta de Twitter del menor para, posteriormente en el desarrollo web, extraer información. Por otro lado, la finalidad del servidor es la extracción y análisis de los datos para después mostrarla en la página web desarrollada. En la Figura 3 se muestra la estructura general de la herramienta, que consta de cuatro módulos principales (aplicación Android, servidor web, aplicación de desarrollo de Twitter y base de datos).

En el diagrama de la Figura 4 se muestra el registro en la base de datos por primera vez de un menor. Una vez obtenidos los valores que identifican al menor tras lanzar la actividad de autenticación y el email del supervisor, se realiza una petición POST al servidor para llevar a cabo la vinculación. El servidor tendrá que comprobar si existe una cuenta de supervisor con ese email y responder con un código y texto a la app android. Por último se actualiza la lista de supervisores mediante otra petición al servidor, en

este caso una petición GET. Cuando el servidor responde con la lista de supervisores, en la aplicación móvil se elimina la lista local y se sustituye por la nueva. El proceso de implementación de la herramienta, así como todos los ficheros desarrollados, se detalla en el memoria completa.

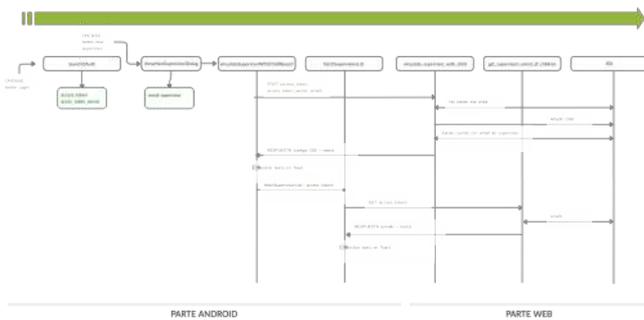


Fig. 4. Registro en la base de datos por primera vez de un niño.

### V. RESULTADOS

Una vez se ha descargado e instalado la aplicación Android (Figura 5) el funcionamiento de la herramienta es el mostrado en la Figura 6.



Fig. 5. App Android.



Fig. 6. Procedimiento general a seguir para el uso de VORDRA.

En la Figura 7 se observa el resultado del análisis realizado. Esta pantalla se muestra al hacer *click* en el identificador de un hijo. Siguiendo el diseño realizado anteriormente, consta de una lista de las palabras más usadas en la esquina superior izquierda. En la derecha se encuentra la gráfica de barras con las horas de *posteo* más frecuentes, y debajo de esta el gráfico de barras con el porcentaje de tweets positivos, negativos o neutros. Por último, en la esquina inferior izquierda se muestra una lista con los usuarios bloqueados. Puede hacerse *click* tanto al usuario

del menor, que se muestra encabezando la página, como a los usuarios bloqueados para acceder a su página de Twitter.

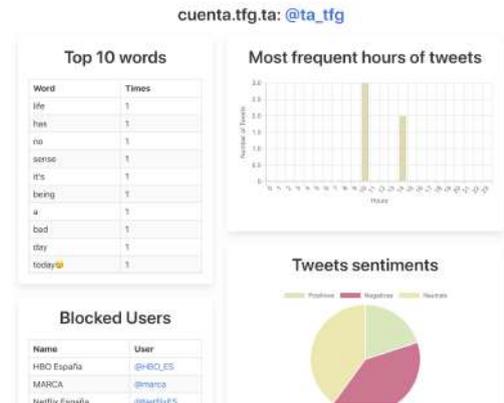


Fig. 7. Resultados mostrados en la página web.

### VI. EVALUACIÓN

Se estableció la finalidad de la herramienta como un método para dotar a los padres de una fuente de información fidedigna sobre el uso de sus hijos/as de Twitter. Sin embargo, la meta final de VORDRA no es la de establecer un diagnóstico como tal. Para poder realizarlo han de tenerse en cuenta numerosos factores que difícilmente pueden conocerse únicamente a través de una red social, y menos si la herramienta que proporciona dicha información pretende ser no invasiva. Es por esto que se cometería un error al intentar evaluar la herramienta implementada en estos términos: si puede detectar casos de adicción, es válida, y en caso contrario no. Deberá evaluarse como una herramienta que sirve de ayuda a los progenitores, quizá a establecer una sospecha que posteriormente puede ser confirmada o desmentida, o quizá a afirmar o refutar una ya implantada.

Como no ha sido posible testar la herramienta en población real, se ha esbozado un escenario, que muestra el comportamiento que un menor con un estado de ánimo depresivo podría tener. Ha sido este el escenario elegido por una razón principal: los distintos problemas que se han mencionado pueden desembocar todos en una misma sintomatología de tipo depresivo. Debe considerarse que esbozar un comportamiento definido conlleva una gran complejidad, ya que cada menor actuará de manera distinta ante una misma situación pues interfieren factores muy diversos.

En la Figura 8 se puede observar el análisis de la cuenta de Twitter de un menor que comienza a sufrir algún problema que afecta a su bienestar emocional. En un principio el *posteo* de *tweets* se realiza en horas normales del día y se detecta un sentimiento positivo en la mayoría de los *tweets*, siendo nula la presencia de sentimientos negativos. En este caso no tendríamos por qué encontrar un gran número de usuarios bloqueados y en las palabras más frecuentes no habría nada que llamase la atención. Finalmente pasamos a la tercera parte del escenario, el momento en que el estado

de ánimo del menor se ve gravemente perjudicado.

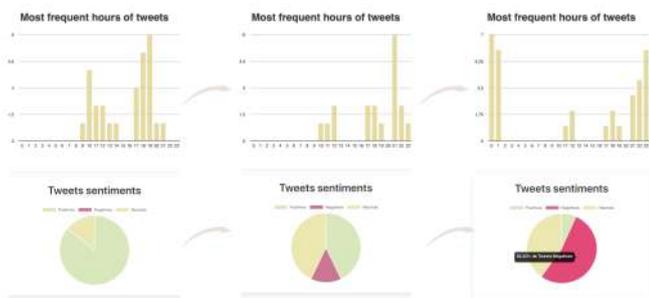


Fig. 8. Escenario ficticio del comportamiento de un menor.

## VII. CONCLUSIONES Y LÍNEAS FUTURAS

En este trabajo se ha propuesto una herramienta innovadora en su modo de presentación al menor, que respeta su intimidad, apostando por la posibilidad del establecimiento de un mutuo acuerdo entre progenitores e hijos. VORDRA facilita a los padres la obtención de una información sobre el uso de la red social Twitter, que resulta altamente valiosa para alertar, confirmar o desmentir sospechas sobre un problema derivado del uso de las redes sociales, tales como adicción, ciberacoso o depresión. Como se mencionó con anterioridad, no se ha establecido como objetivo la realización de un diagnóstico contundente e irrefutable sobre la presencia, o no, de estos problemas.

A nivel personal, puedo asegurar que enfrentarme a la realización de un proyecto de estas características, diseñando y posteriormente implementando una solución a un problema real con un trabajo continuo de documentación, ha sido altamente gratificante. Ha sido necesario un estudio del estado del arte en profundidad, conocer qué problemas específicos se derivan de las redes sociales y cómo se enfrenta a ellos una población tan vulnerable, que carece de los recursos necesarios para afrontar este tipo de situaciones. Asimismo, ha sido indispensable conocer hasta qué punto, de manera legal, pueden los padres acceder a las redes sociales de sus hijos/as y, una vez averiguado, decidir qué grado de “invasividad” sería el más adecuado. En el aspecto técnico, he tenido que conocer y aprender una gran variedad de tecnologías como ha sido programación en Android Studio, manejo de la API de Twitter desde Android, lenguaje HTML, el uso de múltiples librerías, etc. Tecnologías con las que no había experimentado anteriormente.

En este proyecto se han establecido los cimientos de una herramienta que permita asegurar el bienestar emocional y el correcto uso de Twitter en niños y adolescentes, para ello se ha implementado una primera versión que dota a los padres de cierta información de sus hijos/as. Se trata de algo tremendamente necesario en la época en que vivimos, donde los smartphones, y por ende las redes sociales, parecen haber pasado a ser una extensión más en el cuerpo de muchos jóvenes. Sin embargo, dedicando más tiempo a su desarrollo e incluyendo el análisis a un mayor número de parámetros, esta herramienta tiene un amplio margen de mejora. Algunos

de estos avances serían:

- Mayor número de parámetros analizados.
- Sistema de alarmas.
- Análisis histórico.
- Establecimiento de un dominio y servidor no local.
- Evaluación en población real.

En suma, la herramienta VORDRA tiene un posible alcance aún por estudiar enorme.

## AGRADECIMIENTOS

Llegados a este punto, me gustaría mostrar mi gratitud a todas las personas que han hecho posible este proyecto.

Ante todo, quisiera dar las gracias a mi tutor Don Pedro García Teodoro por haberme guiado en todo el proceso. Muchas gracias por haber confiado en mí y por todo el tiempo que me ha dedicado. Sin su ayuda no habría sido posible.

Indudablemente a mi madre, no concibo mejor ejemplo de amor y valentía. Gracias, porque sin ti no tendría nada que poder agradecer. A mi padre, gracias por tu cariño infinito. Y a mi abuela, gracias por tu dulzura y tus consejos.

## REFERENCIAS

- [1] FUNDACIÓN ANAR, ESTUDIO SOBRE EL CIBERBULLYING SEGÚN LOS AFECTADOS. <https://www.anar.org/estudio-ciberbullying/>
- [2] ESTUDIO REDES SOCIALES MÁS UTILIZADAS EN ESPAÑA EN 2019 <https://www.perception.es/noticias/2019/10/23/las-redes-sociales-mas-utilizadas-en-espana-2019>
- [3] ARTÍCULO SOBRE LA NOMOFOBIA Y FOMO <https://desaludpsicologos.es/problemas/adiccion-a-las-tecnologias-tic/nomofobia-y-fomo/>
- [4] DOCUMENTACIÓN API TWITTER [developer.twitter.com](https://developer.twitter.com)
- [5] SITIO WEB OFICIAL DE ANDROID STUDIO [developer.android.com](https://developer.android.com)
- [6] SITIO WEB OFICIAL DE LA LIBRERÍA TWITTER4J [twitter4j.org](https://twitter4j.org)
- [7] DOCUMENTACIÓN DE LA LIBRERÍA RETROFIT [square.github.io/retrofit/](https://square.github.io/retrofit/)
- [8] SITIO WEB OFICIAL DE VISUAL STUDIO [code.visualstudio.com](https://code.visualstudio.com)
- [9] SITIO WEB OFICIAL DE PYTHON [python.com](https://python.com)
- [10] SITIO WEB OFICIAL DE SQLITE [sqlite.org](https://sqlite.org)



**Beatriz Quesada Tobar** nacida en Granada el 29 de Octubre de 1999, graduada en Ingeniería de Tecnologías de Telecomunicación, con mención en Telemática, por la Universidad de Granada en el año 2020. Alumna del Máster de Ingeniería de Telecomunicación impartido por la Escuela Técnica Superior de Ingeniería de Telecomunicación (ETSIT) de la Universidad Politécnica de Madrid (UPM).

# DESARROLLO DE UN CRAWLER PARA ANÁLISIS DE LA CONECTIVIDAD EN LA RED I2P

Autor: Alberto Abellán Galera, e-mail: albertoabellan@correo.ugr.es

Tutor: Gabriel Maciá Fernández, e-mail: gmacia@ugr.es

Titulación: Ingeniería de Tecnologías de Telecomunicaciones

Departamento de Teoría de la Señal, Telemática y Comunicaciones

Universidad de Granada

**Resumen**—El servicio estrella de Internet, la WWW, tiene una cara oculta que muchos usuarios ignoran: las Darknets, redes “anónimas” a las que se accede mediante software específico o configuraciones especiales. La más conocida, TOR, ha sido la más estudiada y analizada, no siendo así el caso de otras como I2P o Freenet.

Este proyecto pretende abordar el problema de explorar e investigar la Darknet I2P desde la perspectiva y enfoque del crawling. Para ello, se llevará a cabo el desarrollo de un crawler que sea capaz de extraer el mayor número posible de eepsites (websites en I2P) así como las relaciones entre ellos, intentando dar a los resultados forma de mapa de interconectividad de eepsites. De esta manera, se buscará extraer conclusiones acerca de la arquitectura de esta red (si tiene la mayoría de eepsites dispersos, si están conectados entre sí y la red es más o menos centralizada, etc).

**Palabras clave**—Anonymous Browsing, Crawler, Crawling, Dark Network, Deep Web, Eepsite, Hidden Network, I2P, Networks, Python, Robot, Scrapy.

## I. INTRODUCCIÓN

### A. Motivación

La mayor parte de usuarios navega dentro de la Surface Web (el contenido de la WWW que puede ser indexado por los buscadores corrientes). Sin embargo, la Surface Web es tan solo la punta del iceberg de la WWW. Según varios estudios (consultar [1] y [2]), se estima que la Surface Web contiene menos del 10% de todo el contenido existente.

La gran parte del contenido de la WWW se aloja en la conocida como Deep Web, que comprende todo aquello que los motores de búsqueda corrientes no son capaces de indexar. Y dentro de la Deep Web, hay unas redes “anónimas”, llamadas Darknets, a las cuales solo se puede acceder realizando configuraciones especiales o utilizando software específico.

Algunas de las Darknets más conocidas son TOR, I2P y Freenet. Pese a que cada una presenta sus propias particularidades todas comparten la misma filosofía: permitir al usuario hacer uso de la WWW de forma anónima, con privacidad y libertad. El anonimato que las Darknets son capaces de brindar, las convierte en “afines” a actividades ilícitas. Es por esto que también las Darknets dan cobijo a los criminales y albergan contenido ilegal y realmente oscuro.

De entre las Darknets, TOR ha sido la más explorada, estudiada y analizada. No es tal el caso de otras como Invisible Internet Project (I2P).

### B. Objetivos

Por tanto, el objetivo principal de este proyecto es el diseño e implementación de una herramienta que permita obtener un mapa de la conectividad que tienen entre sí los eepsites en la Darknet I2P. De manera que se puedan extraer conclusiones acerca de la arquitectura de esta red y obtener información sobre posibles patrones interesantes de conectividad para el desarrollo de futuras aplicaciones o estudios.

Como objetivos específicos cabe destacar:

- 1) Desarrollar un *crawler* para I2P que permita recopilar información sobre la conectividad que tienen entre sí los distintos eepsites.
- 2) Implementar una herramienta que permita visualizar gráficamente la estructura de la Darknet I2P.

### C. Conceptos importantes

**Crawler:** Un *crawler*, también conocido como *spider* o araña web, es un programa que, de manera automática y sistemática, explora sitios web y recorre las páginas que contienen, buscando dentro de ellas cierta información.

**Eepsite:** Los “eepsites” se podría decir que son a la red I2P lo que los websites a la *Surface Web*. Es decir, los eepsites son los sitios web que se encuentran en I2P.

Para conocer cómo funciona I2P en detalle (recomendable para entender correctamente este proyecto) acudir al libro de la referencia [3].

**Respecto a los mapas:** Vamos a obtener mapas similares al que se muestra en la Figura 1, pero a gran escala. Cada nodo (A, B, C, etc) representa un eepsite. La flechas representan enlaces entre eepsites. Por ejemplo, la flecha que apunta desde E hacia B significa que en el código fuente de una de las páginas web del eepsite E, existe un enlace hacia una página del eepsite B. De esta manera, se dice que “E apunta hacia B”. Para cada nodo se definen las siguientes variables:

- outgoing sites: hace referencia al número de eepsites o nodos hacia los que apunta dicho nodo; por ejemplo, para A: 3 (apunta hacia B, D y E).
- incoming sites: hace referencia al número de eepsites que apuntan hacia dicho nodo; por ejemplo, para A: 4 (es apuntado por B, C, D y E).
- degree: se asigna en función de los incoming sites; por ejemplo, para A: 3.

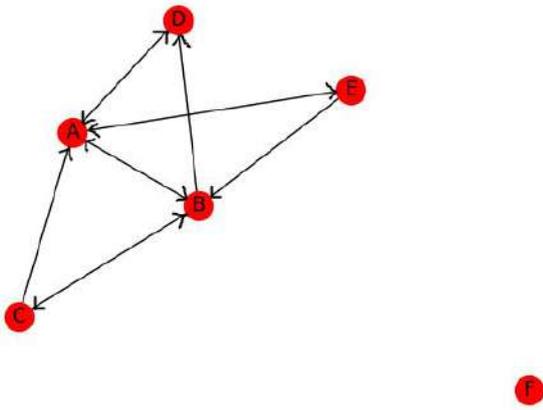


Fig. 1. Ejemplo de mapa de nodos.

## II. PLANIFICACIÓN Y COSTES

Se procede a continuación, de forma breve, a indicar las diferentes fases (y sus tareas) que componen el desarrollo del presente proyecto, así como las fechas en que se llevan a cabo:

- 1) **Fase de planificación (2/10/17 - 8/10/17).**
  - Planificación y definición de requisitos (2/10/17 - 8/10/17)
- 2) **Fase de Estudio y Análisis (9/10/17 - 19/11/17).**
  - Estudio de I2P (9/10/17 - 22/10/17)
  - Estudio del framework Scrapy (23/10/17 - 5/11/17)
  - Estudio de la librería d3.js y el algoritmo K-Shell (6/11/17 - 19/11/17);
- 3) **Fase de Diseño (20/11/17 - 17/12/17).**
  - Diseño del sistema (20/11/17 - 17/12/17)
- 4) **Fase de Implementación (29/01/18 - 8/04/18).**
  - Implementación del *crawler* (29/01/18 - 11/03/18)
  - Implementación de la interfaz de visualización web (12/03/18 - 8/04/18)
- 5) **Fase Experimental (9/04/18 - 20/05/18).**
  - Obtención de las direcciones semilla (9/04/18 - 15/04/18)
  - *Crawling* de la red I2P (16/04/18 - 6/05/18)
  - Procesado de los datos (7/05/18 - 20/05/18)
- 6) **Fase de Documentación (21/05/18 - 24/06/18).**
  - Escritura de la memoria (21/05/18 - 24/06/18)

Finalmente, en la Tabla I se incluye un presupuesto del coste total del desarrollo del proyecto (una vez calculados los diferentes costes de los distintos recursos y recopilando los mismos), el cual tendría un coste total aproximado de 12780,48 €.

Tabla I  
TABLA COSTE TOTAL.

Tipo Recurso	Coste (€)
Software	410,00
Hardware	1635,89
Humano	10800,00
Indirecto	24,59
<b>Total</b>	<b>12870,48</b>

## III. ANÁLISIS

### A. Funcionalidades

Las funcionalidades que se esperan alcanzar son las siguientes:

- El desarrollo de un programa (*spider*) para I2P que permita recopilar la información de los enlaces que contiene un determinado *eepsite*.
- El desarrollo de un programa de control (*crawler*) que lance instancias del *spider* para *crawlear* todos los *eepsites* posibles en la menor cantidad de tiempo posible y que guarde los datos extraídos en una base de datos.
- El desarrollo de una interfaz web que permita visualizar de manera gráfica (en forma de mapa de interconectividad de nodos) la información recolectada por el *crawler* y almacenada en la base de datos.
- Que todo el proceso de *crawling* se realice en el menor tiempo posible y que se actualice la información mostrada al usuario en la interfaz web de manera periódica.

### B. Scrapy

Las principales razones por las que se elige Scrapy de entre todos los posibles frameworks (WebSphinx, JSpyder, WIRE, Pavuk, mnoGoSearch, Web2disk, Heritrix, Apache Nutch, etc) para desarrollar el *crawler* son las siguientes:

- Es uno de los frameworks más utilizados para realizar *web scraping* y *web crawling*.
- Echando un ojo a [4] se puede ver la preocupación por la mejora continua del framework en cuanto a adición de nuevas características y corrección de errores.
- El lenguaje que utiliza es Python, uno de los lenguajes más claros, “directos” y sencillos.
- Tiene una comunidad “vibrante”, basta con echar un vistazo a las miles de preguntas que se recogen en Stack Overflow [5], la mayoría de las cuales reciben respuesta en pocos minutos. Este hecho supone una gran ayuda.

## IV. DISEÑO E IMPLEMENTACIÓN

En este capítulo se procede a resumir la solución propuesta del sistema desarrollado, definiendo la arquitectura del sistema, cada una de las partes que lo forman y sintetizando de qué manera se ha implementado.

La herramienta consta de dos partes principales: el **Crawler** y la **Interfaz**.

- **Crawler:** “Manager”, se encarga de controlar todo el flujo del *crawling*. Lanza instancias de “Spider” (una por cada *eepsite*), las cuales se encargan de recolectar la información en la red I2P. “Manager” procesa esta información recogida y la almacena en la “Database” con el formato adecuado. Además, genera el archivo “data” en formato json que contiene los datos que posibilitan a la otra parte (Interfaz) dibujar el mapa de nodos.
- **Interfaz:** El usuario, a través de su navegador, se conecta a un servidor web que es capaz de servirle páginas web. Este servidor web, de la mano de las tecnologías HTML, CSS y Javascript (front-end) y PHP y SQL (back-end) muestra al usuario en forma de interfaz gráfica (mapa de

interconectividad de nodos) los resultados del proceso de *crawling* llevado a cabo por la parte del Crawler.

En la Figura 2 se muestra el diseño del sistema.

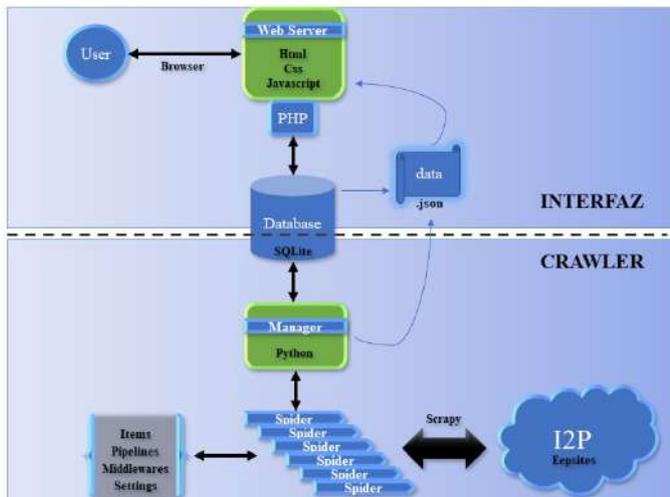


Fig. 2. Diseño del sistema.

### A. Database

Se ha utilizado una base de datos de tipo SQLite, formada por cuatro tablas:

- La tabla Nodes, que representa los eepsites (nodos) y consta de los campos: id, name, outgoing\_sites, incoming\_sites y degree.
- La tabla Links, que representa los enlaces entre eepsites y consta de los campos id, source (eepsite origen) y target (eepsite destino) (contienen la referencia del id del nodo de la Tabla Nodes).
- Las tablas Incoming\_Top y Outgoing\_Top representan el TOP 5 de los eepsites que tienen más incoming sites o más outgoing sites, presentan la misma forma, y contienen la referencia al id del nodo de la tabla Nodes.

### B. Crawler

La estructura de directorios de la parte del Crawler se muestra en la Figura 3.

```

/i2p:
  manager.py
  scrapy.cfg
/i2p:
  __init__.py
  items.py
  settings.py
  pipelines.py
  middlewares.py
/spiders:
  __init__.py
  spider.py
/finished:
/ongoing:

```

Fig. 3. Estructura de directorios de la parte del Crawler.

Para desarrollar esta parte se utiliza el lenguaje Python y el framework Scrapy.

Los ficheros *items*, *pipelines*, *middlewares* y *settings* tienen como misión, entre otras:

- Definir el formato de la información que se va a extraer de capa eepsite, que será una lista de eepsites a los que apunta.
- Priorizar peticiones a recursos que puedan devolver código fuente donde se puedan localizar enlaces hacia eepsites, es decir, no se realizarán peticiones a aquellas URLs con extensiones por ejemplo de imágenes, vídeo o audio para ahorrar en tiempo y aumentar la eficiencia del Crawler.
- Para estar en disposición de alcanzar mejores resultados, se establecen ciertos ajustes como: no respetar las reglas de los ficheros robots.txt, un tiempo de delay de 1 segundo entre peticiones seguidas a páginas de un mismo eepsite y un número máximo de 3 intentos antes de dar por caído un determinado eepsite.

Después, tenemos los dos programas principales:

- El *spider* o araña es el módulo que se encarga de extraer todos los enlaces a eepsites contenidos en un determinado eepsite. Interactúa con los módulos mencionados anteriormente, con la red I2P y con el manager. Crea ficheros .json en el directorio /ongoing donde se guarda el resultado parcial del crawling, además crea ficheros vacíos .ok o .fail en el directorio finished en función de si un eepsite ha sido crawlado correctamente o no, respectivamente.
- El programa más importante, el *manager*, es el núcleo del Crawler y de todo el sistema. Se encarga de “controlar el tempo” de todo el proceso. Lleva un control de los eepsites que han sido visitados, los que están pendientes, los que están siendo visitados y los que han sido descartados. Contiene una lista de URLs semilla por donde comienza el crawling. Lanza una instancia del spider por cada eepsite. Se pueden lanzar simultáneamente varios spiders para ganar en rendimiento (máximo 20, para no saturar los recursos de CPU y memoria del ordenador que se dispone). También procesa los ficheros .ok, .fail y .json generados por los spiders y almacena toda la información extraída en la base de datos. Para dotar al sistema de dinamismo, cada minuto actualiza el fichero i2p\_data.json, fichero que permite a la Interfaz dibujar el mapa de interconectividad de nodos. Además genera un fichero de log donde quedan registradas todas las operaciones que realiza el programa y el estado de las mismas.

### C. Interfaz

Se utiliza el paquete Ampps para crear el servidor web donde se alojarán las dos páginas que mostrarán al usuario los resultados del crawling: index y top. Este servidor web, de la mano de las tecnologías HTML, CSS y Javascript (front-end) y PHP y SQL (back-end) muestra al usuario en forma de interfaz gráfica los resultados del proceso de crawling llevado a cabo por la parte del Crawler.

La estructura de directorios de la parte de la Interfaz se muestra en la Figura 4.

La información se presentará al usuario a través de dos páginas: index y top.

```

/www:
  eepsite.css
  index.html
  top.html
  info.php
  reaction.php
  top.php
  i2p_data.json
  i2p_database.db
/css:
  fondo.png
  brown.png
  ...
/favicon:
  favicon.ico
  android-icon-48x48.png
  ...
/example:
  index.html
  miserables.json
/error:
  50x.html
  404.html
  497.html
/cgi-bin:
  perlinfo.pl
  phpinfo.cgi
    
```

Fig. 4. Estructura de directorios de la parte de la Interfaz.

1) *Página index*: Esta página principal es la que muestra el mapa de interconectividad de nodos, tomando el fichero `i2p_data.json` generado por el manager y ayudándose de la librería `d3.js` para generar el dibujo y las animaciones. El mapa presenta las siguientes particularidades:

- Cada nodo (dibujado como un círculo) representa un *eepsite*.
- Cada link (dibujado como una línea gris) representa un enlace entre dos *eepsites*.
- El tamaño y el color del nodo dependen del número de `incoming_sites` de dicho nodo (a mayor número de `incoming_sites`, mayor tamaño).
- El mapa ofrece una pequeña animación que permite desplazar los nodos a gusto del usuario pinchando sobre los mismos y arrastrando.
- Cuando se pincha sobre un nodo, se muestra en el div “Node Information” la información del nodo (tarea que lleva a cabo el script `reaction.php`).
- Cuando se sitúa el puntero del ratón sobre un nodo, se muestra su nombre. El contenido de esta página principal se actualiza cada minuto (definido así en `manager.py`).

2) *Página top*: La página `top.html` muestra lo siguiente:

- En el div “Total Information” indica cuántos nodos totales se han descubierto así como cuántos nodos de cada grado hay.
- En el div “TOP 5 – Incoming Sites” se muestran los nodos pertenecientes al Top 5 de *eepsites* con mayor número de `incoming_sites`: aquellos *eepsites* dentro de I2P que son más apuntados por otros *eepsites*.
- Por último, en el div “Top 5 – Outgoing Sites” se muestran los nodos pertenecientes al Top 5 de *eepsites* con mayor número de `outgoing_sites` (aquellos *eepsites* dentro de I2P que apuntan a mayor número de *eepsites*).
- Esta página actualiza la información que muestra cada vez que se refresque o recargue. (los scripts `top.php` e `info.php` realizan consultas contra la base de datos).

En las Figuras 5 y 6 se muestra la parte superior de las páginas descritas anteriormente.

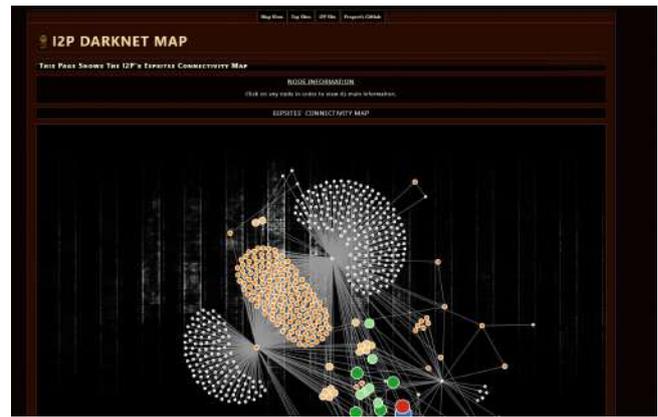


Fig. 5. Página `index.html`



Fig. 6. Página `top.html` (parte superior)

Todo el código desarrollado para implementar el proyecto puede encontrarse en [https://github.com/Abellan09/i2p\\_crawler](https://github.com/Abellan09/i2p_crawler).

## V. EVALUACIÓN Y PRUEBAS

Se parte de las siguientes consideraciones previas:

- Se han realizado 4 pruebas a lo largo de 63 días.
- Cada prueba consiste en poner el marcha el sistema durante un tiempo determinado para obtener resultados. Se ejecuta el programa `manager.py` y se deja corriendo.
- Para todas las pruebas se ha utilizado la misma lista de URLs semilla, formada por un total de 206.
- Cabe resaltar que en ninguna de las pruebas se ha llegado “al final”, es decir, no se ha terminado el crawling de la Darknet I2P, se ha crawlado una parte, lo que el tiempo ha permitido. No se puede estimar el tiempo que se necesitaría para abarcar la red entera.

Tabla II  
TABLA RESUMEN.

Información	Prueba 1	Prueba 2	Prueba 3	Prueba 4
<b>Duración</b>	9 días	14 días	22 días	18 días
<b>Eepsites Totales</b>	609	607	3921	664
<b>Eepsites con Grado 1</b>	398	163	3463	194
<b>Eepsites con Grado 2</b>	196	390	425	398
<b>Eepsites con Grado 3</b>	13	43	25	59
<b>Eepsites con Grado 4</b>	2	8	5	6
<b>Eepsites con Grado 5</b>	0	3	3	5
<b>Eepsites con Grado 6</b>	0	0	0	2

### A. Primera prueba

La primera prueba se llevó a cabo entre los días 25/06/2018 y 03/07/2018; se consiguieron descubrir 609 *eepsites*.



#### D. Cuarta prueba

Finalmente, la cuarta prueba se llevó a cabo entre los días 08/08/2018 y 26/08/2018; se consiguieron descubrir 664 *eepsites* (ver Tabla II).

En el mapa (Figura 9) podemos ver que si bien es cierto que en esta última prueba es en la que encontramos mayor presencia de *eepsites* con grados grandes (nodos de color rojo, rosa y marrón), vuelven a predominar los *eepsites* de tamaño reducido, es decir, aquellos que son apuntados por muy pocos otros *eepsites*, como ya viene siendo tónica en las pruebas anteriores (colores morado y azul en este caso).

En los TOP5 tanto de incoming como de outgoing vuelven a aparecer viejos conocidos de las pruebas anteriores, como el *eepsite* 3mzmrus2oron5.....i2p, que apunta a 426 *eepsites*; o difracker.i2p, siendo en este caso apuntado por 11 *eepsites*.

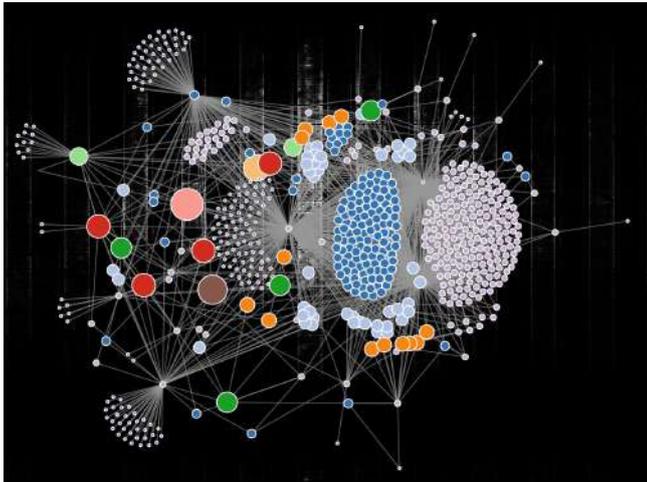


Fig. 10. Mapa. Prueba 4.

## VI. CONCLUSIONES

Se pueden extraer las siguientes **conclusiones**:

- Realizar el *crawling* de I2P es una tarea que requiere de mucho tiempo y recursos computacionales, por lo que se ha conseguido *crawlear* una parte de dicha red.
- Por la forma que toman los mapas obtenidos, se puede observar que la *Darknet* I2P es una red descentralizada, en la que la mayoría de los nodos (*eepsites*) que la conforman son nodos “aislados”, que apuntan o son apuntados por muy pocos *eepsites* (la mayoría eran de grado 1 o 2). Esto concuerda con la filosofía de una *Darknet*, y más aún con la de I2P, que al contrario por ejemplo que TOR, busca ser lo más distribuida posible.
- La mayoría de *eepsites* presentes en el Top 5 de Incoming Sites (aquellos *eepsites* más apuntados o referenciados por otros *eepsites*) no apuntan hacia ningún otro *eepsite* (en su código fuente no hay enlaces hacia otros *eepsites* de la red), hecho que resulta más que curioso.
- Se observa también que la mayoría de los *eepsites* que formaban el Top 5 de Outgoing Sites (aquellos *eepsites* que apuntan a más *eepsites*) tienen la dirección en el formato de base32.

Para finalizar, se incluyen algunas posibles mejoras que se podrían añadir al sistema desarrollado como **trabajo futuro**:

- Implementar un programa que pudiera controlar de manera sincronizada varias ejecuciones simultáneas de *manager.py*, de forma que se pudieran ejecutar procesos de *manager.py* en ordenadores diferentes, lo que aumentaría en gran medida la productividad del Crawler.
- Añadir un módulo (otro *spider*) que permitiera realizar lo mismo que se ha hecho para I2P para otras *Darknets*.
- Aprovechando el módulo anterior, estudiar también la posible interconectividad o interacciones que pudieran existir entre las distintas *Darknets* (por ejemplo, *eepsites* en I2P que apunten a sitios alojados en TOR y viceversa).
- Extraer de cada *eepsite*, aparte de los enlaces hacia otros *eepsites*, otro tipo de información que pudiera ser relevante para llevar a cabo otro tipo de estudios.

## AGRADECIMIENTOS

### INFINITAS GRACIAS:

A Gabriel, por su motivación, sus ideas y sus ánimos.

A Sergio, por su ayuda y preocupación desde la distancia.

A mis amigos, por sus imprescindibles balones de oxígeno.

A mis profesores y compañeros de clase, en especial a Pepe, Cristian, Álex, Francis y Gemma.

A mis padres, Valentín y Goyi, por su entrega, su educación y por creer en mí; mis hermanos, Carmen y Eduardo, por su cariño y complicidad; mi familia, por permanecer en pie cuando todo lo demás se derrumba.

A Lorena, la mujer de mi vida.

A Dios, por llevarme siempre de la mano y guiarme.

## REFERENCIAS

- [1] R. A. Rodríguez-Gómez, G. Maciá-Fernández, and A. Casares-Andrés, “On understanding the existence of a deep torrent,” *IEEE Communications Magazine*, vol. 55, no. 7, pp. 64-69, 2017.
- [2] D. S. Rudesill, J. Caverlee, and D. Sui, “The deep web and the darknet: A look inside the internet’s massive black box,” 2015.
- [3] D. Echeverri Montoya, *Deep Web: TOR, FreeNET & I2P. Privacidad y Anonimato*. 0xWord, 2016.
- [4] Scrapy - *Release notes*. <https://doc.scrapy.org/en/latest/news.html>.
- [5] Stack Overflow - *Scrapy*. <https://stackoverflow.com/questions/tagged/scrapy>.



**Alberto Abellán Galera** recibió su título de Ingeniero Técnico de Tecnologías de Telecomunicación en 2018 en la Universidad de Granada. Actualmente se encuentra cursando el máster de Ingeniería de Telecomunicación, es miembro del grupo Simply5 (de soluciones IoT) y miembro del grupo de hacking ético Hackiit.



**Gabriel Maciá Fernández** recibió su título de Ingeniero de Telecomunicaciones en 1998 en la Universidad de Sevilla. Desarrolló su actividad profesional en compañías como Endesa y Vodafone hasta que en 2005 se incorporó a la Universidad de Granada y obtuvo su doctorado tras presentar su tesis “Low-rate Denial of Service Attacks against servers”. Actualmente es profesor asociado del departamento DTSTC de la Universidad de Granada, miembro del grupo de Seguridad e Ingeniería de Redes (NESG) y líder del grupo de hacking ético Hackiit.

# Estudio y análisis de la conectividad en la Deep Web mediante el desarrollo de nuevas herramientas: c4i2p (Crawling for I2P) en I2P

Autor: Alberto Abellán Galera, e-mail: albertoabellan@ugr.es

Tutor: Gabriel Maciá Fernández, e-mail: gmacia@ugr.es

Tutor: Roberto Magán Carrión, e-mail: rmagan@ugr.es

Titulación: Máster de Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—Gran parte del contenido de la Web se aloja en la conocida como *Deep Web*, que comprende todo aquello que los motores de búsqueda no son capaces de indexar. Dentro de la *Deep Web* destacan las *Darknets*, redes anónimas a las cuales solo se puede acceder con configuraciones especiales o software específico. Entre ellas destaca la red TOR por ser la más explorada y estudiada. No es tal el caso de otras como I2P, que se encuentra explorada y analizada en menor medida. Por tanto, en este trabajo se aborda el problema de explorar e investigar la red I2P, desde el enfoque del *crawling*, con el desarrollo en Python de una herramienta (c4i2p) que permita recopilar información sobre los distintos *eepsites* (sitios web) que conforman esta red. Mediante la correspondiente experimentación sobre escenarios reales se extrae información relevante sobre el contenido de los diferentes *eepsites* analizados así como de la propia estructura de la red I2P en términos de conectividad entre *eepsites*.

**Palabras clave**—Redes, I2P, Anonimato, Deep Web, Darknet, Robot, Crawling, eepsite, Hidden Network, Python.

## I. INTRODUCCIÓN

El servicio por excelencia de Internet, la World Wide Web (WWW) o “Web”, tiene una cara oculta que gran número de usuarios ignora. La mayoría de usuarios navega dentro de la *Surface Web* (que corresponde con el contenido de la WWW que puede ser indexado por los buscadores corrientes). Sin embargo, la gran parte del contenido de la WWW se aloja en la conocida como *Deep Web*, la cual comprende todo el contenido de la WWW que los motores de búsqueda corrientes no son capaces de indexar. Además, dentro de la *Deep Web*, hay unas redes “anónimas”, llamadas *Darknets*, a las cuales solo se puede acceder realizando configuraciones especiales o utilizando un determinado software específico.

Si bien es cierto que no hay que dotar a las *Darknets* de un carácter peyorativo, no se puede dejar de lado que son la parte más escondida de la WWW, y donde el nivel de privacidad y anonimato es, sin duda, más alto. Este hecho las convierte en unas redes muy interesantes, por las implicaciones que tiene un nivel muy alto de anonimato. De hecho, están experimentando un gran auge, y el uso de las mismas crece cada vez más. La más sonada, la red The Onion Router (TOR) [1], es de entre todas las *Darknets* la que ha sido más explorada, estudiada y analizada. No es así el caso de otras *Darknets* como Invisible Internet Project (I2P) [2], la cual se encuentra explorada y analizada en menor medida.

Este proyecto aborda el problema de explorar e investigar la *Darknet* I2P desde la perspectiva y el enfoque del *crawling*. Para ello se desarrolla una herramienta de *crawling* a la que he llamado **c4i2p** (crawling for I2P) de código abierto y que se encuentra disponible en [3]. Dicha herramienta es capaz de encontrar y extraer información tanto del contenido de *eepsites* como de las relaciones entre ellos que determinan en cierta medida la estructura de la propia red.

La estructura del documento es la que sigue: Los fundamentos de I2P serán descritos brevemente en la Sección II. Una introducción al estado del arte se abordará en la Sección III. En la Sección IV se describirá en detalle la herramienta implementada (**c4i2p**). Se detallará en la Sección V el entorno experimental. Los resultados y su discusión se presentarán en la Sección VI. Finalmente, en la Sección VII se resumirán las principales conclusiones extraídas.

## II. FUNDAMENTOS DE I2P

Se recomienda consultar con detalle en la página oficial de I2P [2] y el libro [4] la arquitectura, el cifrado y otros conceptos y elementos que dan forma a I2P. Se describen a continuación algunos elementos y nociones básicas:

- **Eepsites**: Se podría decir que son a la red I2P lo que los websites a la *Surface Web*. Es decir, los eepsites son los sitios web que se encuentran en I2P.
- **NetDB**: La base de datos de la red o NetDB es una tabla hash distribuida (Distributed Hash Table - DHT) con una estructura basada en el algoritmo Kademlia, la cual contiene toda la información relativa para que los nodos de la red puedan comunicarse entre sí (gracias principalmente a los mensajes *routerInfo* y *leaseSet*).
- **Nodos Floodfill**: La NetDB se distribuye con una técnica llamada “*Floodfill*” (inundado), la cual se basa en el uso de una serie de enrutadores llamados también nodos *Floodfill*, los cuales se encargan de mantener la base de datos distribuida y de sincronizar la información correspondiente de los *routerInfo* y los *leaseSet* que se encuentran en la red. A diferencia de las autoridades de directorio de TOR, los *Floodfill* que conforman la NetDB no son fijos ni de confianza; cualquier enrutador de la red puede convertirse en un *Floodfill* si está configurado para compartir un gran ancho de banda.

### III. ESTADO DEL ARTE

Desde que de la mano de Tim Berners-Lee surge la WWW a finales de los 80, muchos son los estudios e investigaciones que se han llevado a cabo sobre la misma. Teniendo en cuenta que este servicio está en constante evolución, cada poco tiempo surgen nuevos interrogantes y nuevas cuestiones que merece la pena considerar.

Los artículos más relevantes hasta la fecha sobre el análisis de las darknets (en especial de TOR e I2P), de su contenido y su conectividad son los que se describen a continuación:

En [5] se describió un enfoque para enumerar todos los servicios ocultos (o *hidden services*) en TOR. El análisis de la popularidad de los servicios ocultos de TOR se llevó a cabo en trabajos como el [6].

En 2018, en [7] se lleva a cabo la creación de una base de datos llamada DUTA, que contiene un listado de multitud de *hidden services* de TOR clasificados por contenido. En 2019, continuando con la línea del artículo anterior, se propone en [8] un nuevo algoritmo llamado TORank.

En [9] se intenta realizar un descubrimiento y análisis de los *eepsites* que conforman la red I2P, de manera similar a lo que se hizo para TOR en [7] y [8].

Finalmente, [10] es quizá uno de los estudios académicos sobre I2P más completos hasta la fecha. Los autores encuentran que en aquel momento había alrededor de 32000 pares I2P activos en la red diariamente, y que de estos, 14000 se encontraban detrás de Network Address Translators (NATs) o *firewalls*.

### IV. HERRAMIENTA C4I2P (CRAWLING FOR I2P)

En la presente Sección se procede a describir los principales actores, procedimientos y elementos que dan forma a la herramienta desarrollada: **c4i2p** (Crawling for I2P).

#### A. Actores

De manera simplificada, el diseño de la solución propuesta trata de ofrecer al usuario una herramienta o sistema que le permita “interaccionar” con la red I2P, entendiendo esta interacción en la forma de realizar un *crawling* sobre los *eepsites* que componen dicha red, permitiendo recopilar diferentes datos sobre los mismos. De esta forma (ver Fig. 1), los principales agentes involucrados son el usuario, la propia herramienta **c4i2p** y los *eepsites* que contiene la *Darknet* I2P.

#### B. Procedimientos

La herramienta consta de cuatro **procedimientos** o partes principales inter-relacionadas entre sí:

- El procedimiento de **Management**, que será el gestor principal, el que controlará todo el flujo de trabajo de la herramienta. Regirá la interacción entre el resto de procedimientos y elementos que componen el sistema.
- El procedimiento de **Crawling**, que será el encargado de llevar a cabo la tarea del *crawling* como tal y todo lo que de este subyace para la búsqueda y extracción de la información objetivo de los *eepsites*.
- El procedimiento de **Discovering**, que se encarga de comprobar la disponibilidad de los *eepsites* contemplados

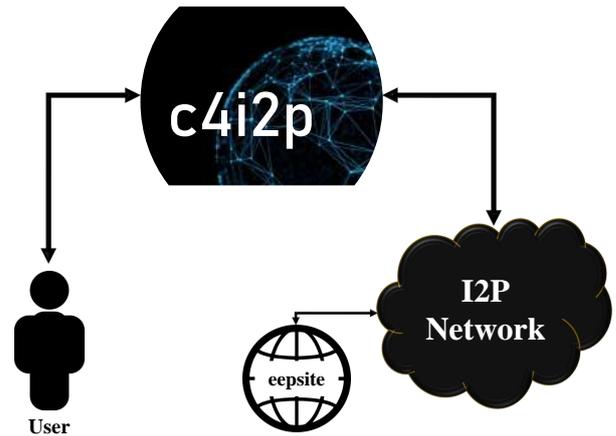


Fig. 1. Actores del sistema.

en el sistema a lo largo del tiempo de forma controlada en el tiempo.

- El procedimiento de **Data Storage**, que es un procedimiento paralelo a todos los demás y del que todos hacen uso, se encarga del almacenamiento, persistencia y gestión de toda la información proporcionada por el resto de procedimientos.

#### C. Elementos y relaciones

Los principales elementos que forman la herramienta **c4i2p** son (ver Fig. 2):

- **Json Files**: Son los ficheros generados por los Spiders que contienen la información extraída de los *eepsites*.
- **Sources**: Las Sources son las direcciones que se corresponden con los nombres de *eepsites* en I2P. Pueden tener tres orígenes diferentes:
  - SEED: pertenecer a la lista de semillas iniciales desde las que parte el experimento.
  - FLOODFILL: aquellos descubiertos por los nodos *Floodfill* desplegados.
  - DISCOVERED: aquellos que han sido descubiertos mediante el proceso de *crawling* (es decir, están enlazados en el código fuente de algún *eepsite*).
- **Manager**: Íntimamente relacionado con el procedimiento de Management, el Manager se constituye como cerebro y núcleo de toda la herramienta. Se encarga de las principales funciones de coordinación del resto de elementos. Lleva un control de los *eepsites* que están activos, de los que no, de los que están pendientes de ser *crawleados*, de los que han finalizado, de los descartados. Se encarga de lanzar los módulos de descubrimiento (Discoverys) y de *crawling* (Spiders), de gestionar las Sources, de volcar la información de los Json Files en la base de datos, de generar un *log* de todas las operaciones, etc.
- **Discovery**: El elemento Discovery implementa la principal funcionalidad del procedimiento de Discovering. Se encarga de comprobar si un *eepsite* se encuentra activo, es decir, si está online y operativo. Se lanza previo a cualquier Spider.

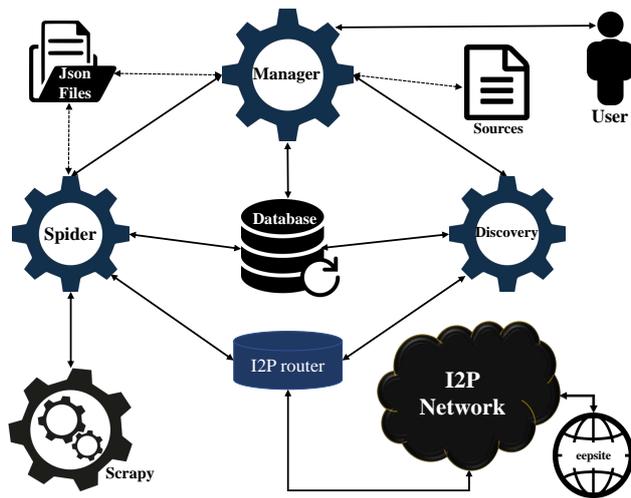


Fig. 2. Diseño del sistema: actores, elementos y relaciones.

- **Spider:** El Spider es el encargado de llevar a cabo el *crawling* como tal de un determinado *eepsite*. De manera recursiva “se sumerge” en todas las páginas del mismo para extraer la información que se desea, entre ella el idioma, los enlaces que apuntan a otros *eepsites*, los scripts e imágenes, una muestra de texto, etc.
- **Scrapy:** El módulo Scrapy facilita enormemente al Spider algunas de sus labores, como por ejemplo establecer el *proxy* o ignorar las instrucciones del *robots.txt*.
- **Database:** La base de datos almacena de forma ordenada y estructurada toda la información extraída de los *eepsites* así como la inherente al propio sistema.
- **I2P Network:** La Red I2P es donde se encuentran alojados los *eepsites*, los cuales son los actores de los que queremos extraer información.
- **I2P router:** Y por último, el I2P router es el elemento que precisamente servirá para conectar los módulos Discovery y Spider con la red I2P.

#### D. Estados de un *eepsite* en *c4i2p*

Por último, se hace necesario definir una serie de estados en los cuales puede encontrarse un determinado *eepsite* cuando está siendo manejado por la herramienta *c4i2p* (el diagrama de flujo entre estados puede consultarse en la Fig. 3):

- **DISCOVERING:** es el primer estado en que un *eepsite* puede encontrarse, representa el momento en que está siendo “descubierto”, es decir, se le están enviando peticiones HTTP para comprobar si se encuentra activo.
- **DISCARDED:** si un nodo se ha intentado descubrir un número máximo de veces sin éxito, se “descarta”, pasando a este estado final, DISCARDED.
- **PENDING:** un nodo se encuentra en este estado una vez se ha comprobado que está activo, es decir, queda “pendiente” o a la espera de intentar ser *crawleado*.
- **ONGOING:** cuando un nodo está ONGOING es porque un Spider se está encargando de *crawlearlo*.
- **ERROR:** un nodo se encuentra en el estado de ERROR si ha habido algún error mientras se estaba *crawleando*.
- **FINISHED:** por último, un nodo se etiqueta como FINISHED si ha terminado de ser *crawleado* correctamente.

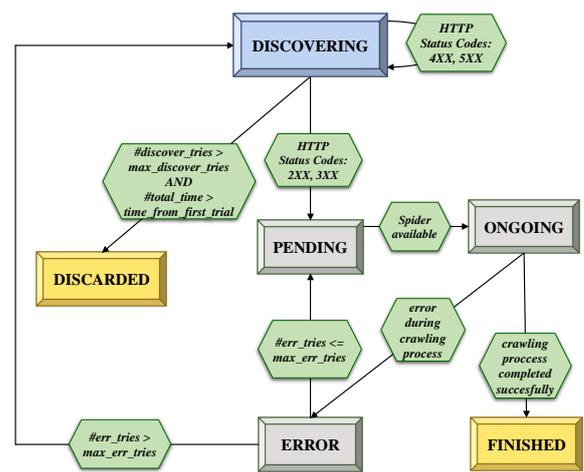
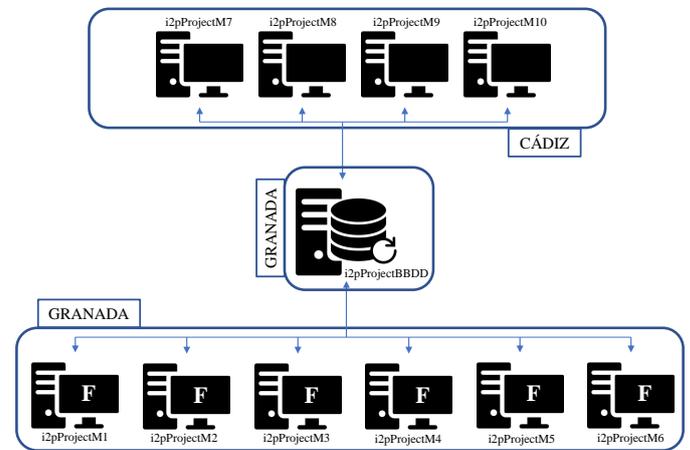
Fig. 3. Diagrama del ciclo de vida (estados) de un *site*.

Fig. 4. Entorno experimental desplegado.

## V. ENTORNO DE EXPERIMENTACIÓN

Para aprovechar todo el potencial de la herramienta, se ha lanzado simultáneamente en varias máquinas de manera distribuida. Así, se han utilizado (ver Fig. 4) un total de 11 máquinas, estando 7 de las mismas localizadas en Granada (dependientes de la UGR) y las otras 4 restantes situadas en Cádiz (dependientes de la UCA). Inicialmente, las *seeds* o semillas se reparten equitativamente entre cada una de las máquinas participantes en el experimento, para que tengan el mismo número de *seeds* por donde empezar. En concreto, 10 de las máquinas correrán *c4i2p* y la otra albergará la base de datos. Su nomenclatura será la siguiente:

- *i2pProjectMX*: máquinas que corren *c4i2p*, donde X es un número entre 1 y 10 que identifica la máquina.
- *i2pProjectBBDD*: máquina que contiene la base de datos.

Las 10 máquinas *i2pProjectMX* se comunicarán con la base de datos *i2pProjectBBDD*, de manera que gracias a esta se consiga mantener la persistencia y consistencia de los datos, para que cada máquina se encargue de “una parte del trabajo”.

Las 6 máquinas en Granada se distinguen de las 4 de Cádiz en la Fig. 4 por la F, que denota “Floodfill”. Esto quiere decir que en las máquinas de Granada los I2P routers se han

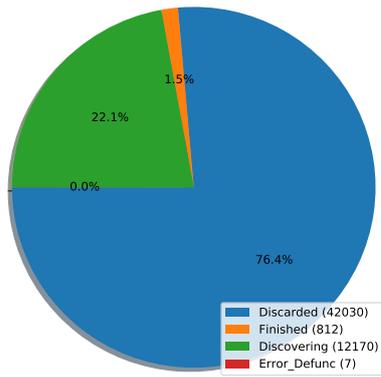


Fig. 5. Sites por estado al finalizar la prueba.

configurado en modo *Floodfill*, lo que permite capturar tráfico de la propia red I2P que facilita descubrir nuevos *eepsites* que añadir al sistema para que procese.

## VI. RESULTADOS Y DISCUSIÓN

En total, durante todo el experimento, el cual ha durado 111 días, la herramienta ha manejado un total de 55019 *eepsites*. De los cuales, tan solo 812 han sido *crawleados* con éxito. La gran mayoría de *sites* *crawleados* han sido descubiertos gracias a los routers I2P desplegados como *Floodfill* (fuente FLOODFILL, el 74.1%).

En el momento de detener el experimento, algo más del 75% de *sites* (42030) habían sido descartados por encontrarse inactivos u *offline* (estado DISCARDED). El 22,1% estaban siendo descubiertos (en el estado DISCOVERING). Apenas el 1,5% de *sites* habían terminado de ser *crawleados* correctamente (812 *sites* en el estado FINISHED). Y finalmente, los 7 que se encontraban en el estado ERROR\_DEFUNC corresponden a los 7 *sites* que estaban siendo *crawleados* en ese momento. Aparecen en este estado en vez de en ONGOING porque los Spiders que los *crawleaban* fueron detenidos bruscamente. Ver Fig. 5.

Tras el comienzo vertiginoso del primer día, donde se *crawlean* 278 *sites* (ver Fig. 6) apreciamos que, como cabría esperar de forma intuitiva, el nº de *sites* *crawleados* aumenta de forma lineal a medida que avanza el experimento en el tiempo. En la Fig. 7 podemos ver cómo el primer día se *crawlean* en total 278 *sites*, siendo 151 de fuente SEED, 57 de fuente DISCOVERED y 70 de fuente FLOODFILL. Podemos ver que, a partir del primer día prácticamente solo se *crawlean* *sites* que hayan sido descubiertos mediante la fuente FLOODFILL.

Se ha podido apreciar también que la gran mayoría de *sites* que han podido ser *crawleados* han necesitado poco tiempo y pocos intentos de descubrimiento.

Se han utilizado dos *engines* distintos para detectar el idioma de los *sites*: Natural Language Toolkit (NLTK) y Google Translator. Presentan discrepancias notorias en sus resultados (ver Tabla I), con lo que la detección del idioma es uno de los puntos a mejorar de la herramienta. Si bien,

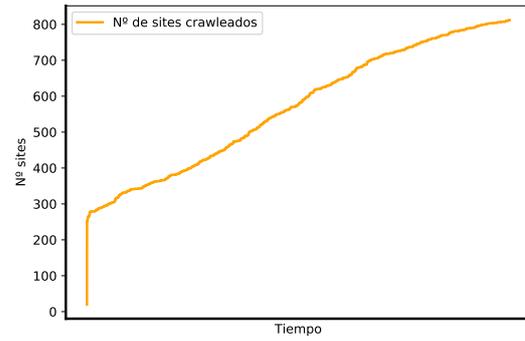


Fig. 6. Evolución temporal del nº de *sites* *crawleados*.

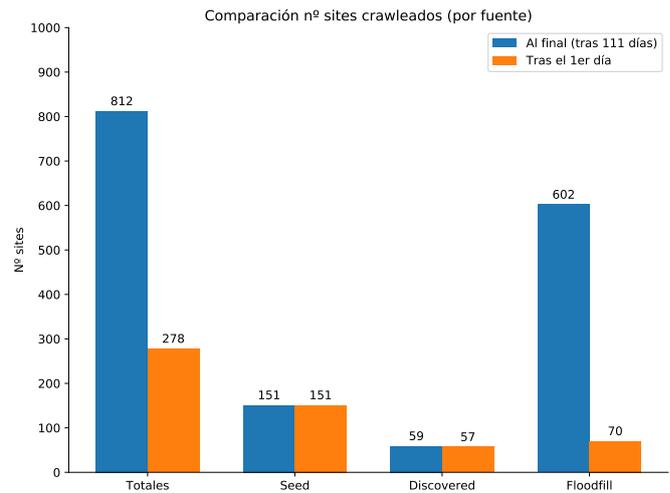


Fig. 7. Comparación del nº de *sites* *crawleados* por tipo de fuente.

podemos apreciar que el idioma predominante en los *eepsites* es el inglés, según ambos *engines*.

Tabla I  
IDIOMA DE LOS *sites* (GOOGLE Y NLTK *engines*).

Idioma	Google engine (%)	NLTK engine (%)
English	96,31	31,04
French	0,86	0,62
German	0,86	0,86
Spanish	0,62	0,99
Norwegian	0,25	1,97
Latin	0,25	0
Italian	0,25	1,23
Welsh	0,12	0
Turkish	0,12	1,60
Portuguese	0,12	7,88
Dutch	0,12	0,62
Hungarian / Russian	0	25,37
Undefined / Other	0,12	27,83

Además, se ha podido apreciar que la mayoría de *sites* apenas contienen texto en su página principal (cuentan con pocas letras y con pocas palabras). Tampoco contienen apenas scripts ni imágenes.

### A. Análisis de la conectividad en I2P

Se procede ahora al estudio de los resultados sobre la característica “estrella” o más importante que se ha considerado:

Tabla II  
TOP 10 NODOS SOURCE: *sites* CON MAYOR *outgoing*.

Site	Intentos de descubrimiento	Duración (min)	Source	Outgoing
Top1out	4	230,80	SEED	399
Top2out	1	45,50	SEED	364
Top3out	1	973,75	SEED	297
Top4out	5	1023,78	SEED	277
Top5out	5	264,43	SEED	172
Top6out	1	30,50	SEED	117
Top7out	14	27121,35	FLOODFILL	78
Top8out	4	249,05	SEED	62
Top9out	2	227,60	SEED	58
Top10out	23	2169,95	FLOODFILL	50

la conectividad entre los *eepsites* que componen I2P.

Entendiendo los *eepsites* como nodos y las relaciones entre ellos como *links* o enlaces, decimos que un nodo apunta a otro si en el código fuente del primero, existe un enlace hacia alguna de las páginas del segundo. Así, se han utilizado dos parámetros principales para el estudio de la conectividad:

- *outgoing*: hace referencia al número de *eepsites* o nodos hacia los que apunta un nodo.
- *incoming*: hace referencia al número de *eepsites* que apuntan hacia dicho nodo.

Estudiando los histogramas se ha apreciado cómo la mayoría de *sites* apuntan hacia muy pocos *sites* (su número de *outgoing* es bajo) y además, se ha podido ver que la mayoría de *sites* son apuntados también por pocos *sites* (su número de *incoming* es bajo).

Sin embargo, a pesar de que la mayoría de *sites* tienen un número bajo de *outgoing*, existen algunas excepciones. Son lo que llamamos **Nodos Source** o Nodos Fuente. El Top 10 se muestra en la Tabla II. Sus conexiones gráficas se muestran en la Fig. 8. Podemos apreciar que el Top 1 apunta a un total de otros 399 *sites*. Es el *site* *3mz...i2p*, que corresponde con el *eepsite* cuyo título es “*eepstatus*”, el cual es un repositorio que contiene multitud de *links* hacia *eepsites*, mostrando además para cada uno cuál fue el último momento en que se comprobó que se encontraba activo. Los sitios añadidos al repositorio más recientemente se muestran al principio de la página.

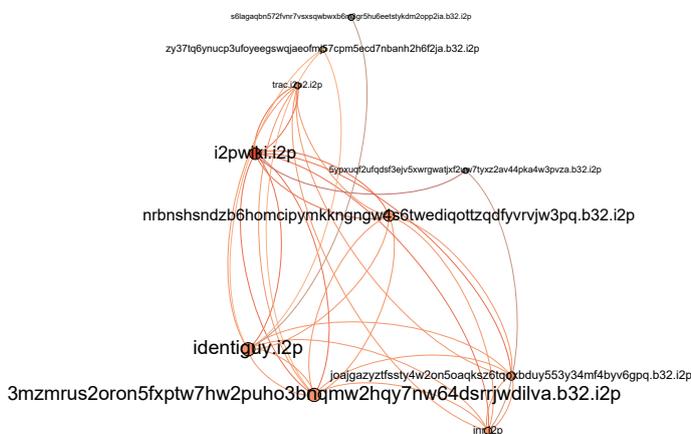


Fig. 8. Gráfico de conectividad entre los TOP 10 Nodos Source.

Por su parte, los *sites* correspondientes al Top 10 de

**Nodos Sink** o Nodos Sumidero, es decir, los *sites* que son más apuntados por otros *sites* se muestran en la Tabla III. Sus interconexiones se muestran en la Fig. 9. Apreciamos que el Top 1 es *stats.i2p*, estando apuntado por un total de 72 *eepsites*. Este *eepsite* muestra estadísticas sobre la red I2P en general y sobre la NetDB en particular. Una diferencia significativa es que los Nodos Sink suelen estar identificados por su dirección en formato legible, al contrario que los Nodos Source, cuya dirección está en formato base 32.

Tabla III  
TOP 10 NODOS SINK *sites* CON MAYOR *incoming*.

Site	Intentos de descubrimiento	Duración (min)	Source	Incoming
Top1in	2	105,68	SEED	72
Top2in	1	503,18	SEED	58
Top3in	5	1023,78	SEED	47
Top4in	1	45,50	SEED	37
Top5in	3	1150,33	SEED	35
Top6in	4	319,07	SEED	32
Top7in	3	179,52	SEED	29
Top8in	3	234,50	SEED	23
Top9in	59	3821,28	SEED	22
Top10in	2	227,60	SEED	21

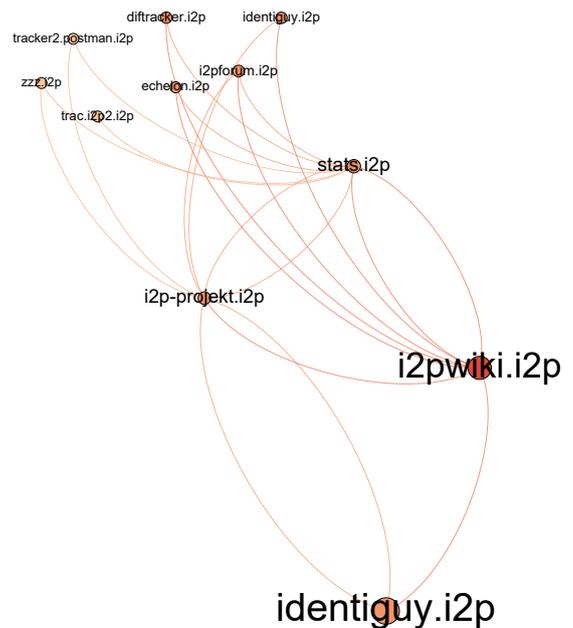


Fig. 9. Gráfico de conectividad entre los *sites* TOP 10 *incoming*.

Para terminar con los resultados, analizamos la distribución de conectividad de los *eepsites* (Ver Fig. 10):

El número total de *sites* que se encuentran totalmente aislados es 517, lo que supone que el 63.67% de *eepsites* que se han podido *crawlear* de I2P son *eepsites* completamente aislados.

El número total de *sites* que no se encuentran aislados o, dicho de otro modo, que se encuentran conectados, es 295; con lo que el 36.33% de *eepsites* que se han podido *crawlear* (franjas verde y naranja) tienen algún tipo de conexión (ya sea *incoming*, *outgoing*, o las dos).

Dentro de estos *sites* no aislados, podemos diferenciar los que están completamente conectados, es decir, aquellos *sites*

que apuntan al menos a otro *site* y que además son apuntados por al menos un *site*. Estos suman un total de 115 *eepsites*, lo que supone un 14,2%.

Distribución conectividad en I2P

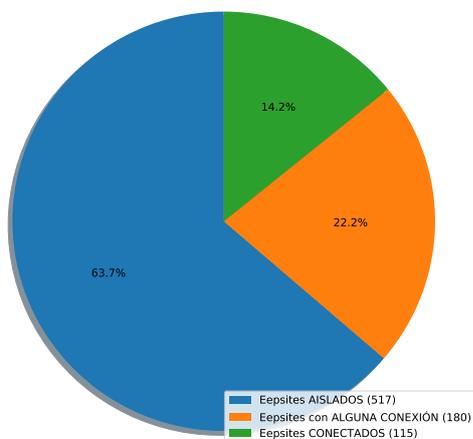


Fig. 10. Distribución de la conectividad de los *sites* *crawleados* en I2P.

## VII. CONCLUSIONES

Tras la realización del presente proyecto, se han podido extraer las siguientes conclusiones:

- Realizar el *crawling* de una red de las características que presenta I2P es una tarea ardua y que requiere de mucho tiempo y recursos computacionales, por lo que se ha conseguido *crawlear* una parte de la red, sin poder determinar con exactitud el porcentaje de la misma.
- La mayoría de *eepsites* que forman el *Top 10 de Nodos Source* (los *eepsites* que apuntan a más *eepsites*) tienen la dirección en el formato de base32, que resulta mucho más difícil de recordar por lo ilegible que es. Contrario a los *sites* presentes en el *Top 10 de Nodos Sink*, donde no se encuentra ninguno en su forma de base32.
- El idioma predominante es el inglés.
- La mayoría de los *eepsites* que se han podido *crawlear* en la red I2P son *sites* “pequeños y sencillos”, esto es, tienen pocas páginas y su página principal incluye poco texto, pocas imágenes y pocos scripts.
- De los *sites* *crawleados*, se aprecia que aproximadamente la mitad (52.09%) se encuentran disponibles (*online*) prácticamente todo el tiempo, mientras que la otra mitad se encuentran activos de manera intermitente (han necesitado más de 5 intentos de descubrimiento).
- La *Darknet* I2P es descentralizada. La mayoría de los nodos (*eepsites*) que la conforman son nodos “aislados”, que apuntan o son apuntados por muy pocos *eepsites*. Esto concuerda con la filosofía de una red oculta o *Darknet*, y más aún con la de I2P, que al contrario por ejemplo que TOR, busca ser lo más distribuida posible.

## AGRADECIMIENTOS

A Gabri y Rober, por sus consejos, ayuda y paciencia.  
A mis amigos, por sus imprescindibles balones de oxígeno.

A mi familia, en especial mis padres, Valentín y Goyi, por su entrega, su educación y por creer en mí; y mis hermanos, Carmen y Eduardo, por su cariño y complicidad.

A Lorena, la mujer de mi vida.

A Dios, por llevarme siempre de la mano y guiarme.

## REFERENCIAS

- [1] TOR Official Site: <https://www.torproject.org>.
- [2] I2P Official Site: <https://geti2p.net/es>.
- [3] GitHub Repository - *c4i2p*: [https://github.com/nesg-ugr/i2p\\_crawler](https://github.com/nesg-ugr/i2p_crawler)
- [4] D. Echeverri Montoya, *Deep Web: TOR, FreeNET & I2P. Privacidad y Anonimato*. OxWord, 2016.
- [5] A. Biryukov, I. Pustogarov, and R.-P. Weinmann, “Trawling for torhidden services: Detection, measurement, deanonymization,” in 2013 IEEE Symposium on Security and Privacy. IEEE, 2013, pp. 80–94.
- [6] A. Biryukov, I. Pustogarov, F. Thill, and R.-P. Weinmann, “Contentand popularity analysis of tor hidden services,” in 2014 IEEE 34th. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, 2017, pp. 35–43.
- [7] M. W. Al Nabki, E. Fidalgo, E. Alegre, and I. de Paz, “Classifying illegal activities on tor network based on web textual contents,” in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, 2017, pp. 35–43
- [8] M. W. Al-Nabki, E. Fidalgo, E. Alegre, and L. Fernandez-Robles, “Torank: Identifying the most influential suspicious domains in the tor network,” Expert Systems with Applications, vol. 123.
- [9] Y. Gao, Q. Tan, J. Shi, X. Wang, and M. Chen, “Large-scale discovery and empirical analysis for i2p eepsites,” in 2017 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2017, pp. 444–449.
- [10] N.P. Hoang, P.Kintis, M.Antonakakis, and M.Polychronakis, “An empirical study of the i2p anonymity network and its censorship resistance,” in Proceedings of the Internet Measurement Conference 2018.



**Alberto Abellán Galera** recibió su título de Ingeniero Técnico de Tecnologías de Telecomunicación en 2018 en la Universidad de Granada. Asimismo, en el presente año 2020 ha completado sus estudios del máster de Ingeniería de Telecomunicación. También es miembro del grupo Simply5 (de soluciones IoT) y colabora con el grupo de Seguridad e Ingeniería de Redes (NESG).



**Gabriel Maciá Fernández** recibió su título de Ingeniero de Telecomunicaciones en 1998 en la Universidad de Sevilla. Desarrolló su actividad profesional en compañías como Endesa y Vodafone hasta que en 2005 se incorporó a la Universidad de Granada. Actualmente es profesor asociado del departamento DTSTC de la Universidad de Granada y miembro del grupo de Seguridad e Ingeniería de Redes (NESG).



**Roberto Magán Carrión** es Ingeniero de Telecomunicación por la Universidad de Málaga desde 2008. Más recientemente, en 2016, recibió el grado de Doctor por la Universidad de Granada con mención Cum Laude. Actualmente es profesor ayudante doctor en el departamento de Teoría de la Señal, Telemática y Comunicaciones y miembro del grupo de investigación NESG (Network Engineering Security Group).

# Implementación de un sniffer de información confidencial escrita en un navegador mediante técnicas de hooking

Autor: José Miguel Noguera Comino, e-mail: jmnoguera@correo.ugr.es

Tutor: Gabriel Maciá Fernández, e-mail: gmacia@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—El uso de la web está muy presente en el día a día, con ella se llevan a cabo gran cantidad de transacciones de información, desde redes sociales a movimiento bancarios. Todas estas comunicaciones están respaldadas en su mayoría por HTTPS, un protocolo que dota de seguridad a la comunicación y hace que los parámetros transmitidos viajen seguros entre los miembros de esta. Esta barrera de seguridad hace que la privacidad de la comunicación sea difícil de violar, pero si se ataca a otros instantes de esta donde HTTPS aun no entra en juego el panorama es muy distinto.

El presente proyecto pretende probar si el uso de ataques con técnicas de hooking permiten obtener credenciales antes de que estas sean protegidas por HTTPS. Por ello se lleva a cabo la realización de una prueba de concepto que permita juzgar la viabilidad y capacidades de estas técnicas.

**Palabras clave**—hooking, sniffer, depuración, reversing, pydbg, firefox

## I. INTRODUCCIÓN

EN la actualidad el malware orientado a recabar las credenciales de los usuarios hace uso de técnicas diversas, desde phishing e ingeniería social hasta explotación de vulnerabilidades. Pero todos ellos buscan un punto débil en la comunicación para cumplir su cometido. Cuando el objetivo son comunicaciones HTTPS el enfoque suele ir orientado a atacar al usuario y/o su máquina en distintos grados.

Las técnicas de hooking se categorizan bajo las que atacan al equipo, aunque para ello no hacen uso de mecanismos ilícitos, ya que el hooking es una capacidad derivada de la depuración de programas y juega un papel importante en el desarrollo de software legítimo. Estas técnicas se basan en la capacidad que tienen los depuradores de detener la ejecución de un programa para, teniendo en cuenta la llamada a una función, poder recuperar el valor de los parámetros usados por esta antes o después de su ejecución y manipularlos.

El proyecto desarrollado muestra cómo estas técnicas pueden ser utilizadas para la obtención de información confidencial.

## II. ESPECIFICACIÓN DE REQUISITOS

Para la correcta realización de la prueba de concepto es necesario acotar la funcionalidad esperada y el material utilizado para su implementación.

Se parte de la existencia de un ataque similar recogido en el libro Gray Hat Python [1] hacia el navegador Firefox en

un entorno Windows. Dado el éxito de esta implementación se apuesta por trasladar la idea al software actual, haciéndose uso de Firefox en la rama ESR para 32 bits ejecutándose sobre Windows 10. La elección de la rama ESR viene dada por su mayor consistencia a lo largo del tiempo, haciendo más viable el desarrollo en un proyecto largo.

El proceso asociado a la depuración y colocación de los hooks ha de venir dado por un depurador compatible con Python y que cuente con capacidades de hooking. Entre las opciones aparecen PyDbg, WinAppDbg, PyKd y Radare2, se opta por el primero ya que presenta una menor curva de aprendizaje y cumple con los requisitos pedidos.

Para facilitar el desarrollo, se complementa a PyDbg con un depurador dinámico con interfaz gráfica, lo cual hace que el estudio del software depurado y su caracterización sea más rápida y simple.

La implementación final ha de corresponderse con un software autónomo y silencioso, que no precise de software previo para su correcto funcionamiento. Para conseguir esto se requiere de una aplicación que convierta un proyecto en Python en un ejecutable. La principal opción era py2exe, pero se optó por PyInstaller en pos de su mayor portabilidad al incluir las DLL's requeridas.

Dado que el punto de hooking utilizado puede variar, se especifica que el código asociado a él permita definir nuevos puntos de hooking con facilidad.

El esquema de comunicación entre el usuario infectado y el equipo que almacena los datos obtenidos, se implementa como un comunicación varios clientes - un servidor. Un esquema verosímil para un software que ha de centralizar la información de varios equipos. A nivel de protocolo de comunicación se elige un desarrollo propio sobre TCP basado en mensajes y con cifrado. Se podría haber optado por un desarrollo más complejo que diese respuesta a más problemáticas, pero dado el enfoque del proyecto no se relega en esta parte una gran carga.

## III. DISEÑO

### A. División funcional

Para realizar el diseño se ha optado por una división funcional que separa todo el código en pequeñas partes, siendo cada una de estas la solución a un problema concreto y la suma de todas la que posibilita la prueba de concepto.

Las divisiones definidas son:

- Hooking y recogida de información
- Gestor del cliente de la comunicación
- Gestor del servidor de la comunicación
- Protocolo de comunicación
- Gestor de la base de datos
- Instalador/Downloader

Estas divisiones se han reflejado a la hora de programar en los distintos módulos que componen el proyecto y cuentan con las clases necesarias para resolver su función. Las partes y relaciones entre ellas pueden verse en la Fig. 1.

A continuación se trata cada una de estas divisiones de forma aislada.

1) *Hooking y recogida de información:* Este módulo tiene como objetivo implementar todo lo relativo a la obtención de las credenciales y su procesado. Esto lo consigue haciendo uso de una función *attach* para asociarse al proceso de Firefox, una etapa de procesado de la información que mediante expresiones regulares devuelve una tupla con formato (*site, user, password*) y una cola FIFO que almacena las tuplas extraídas y sirve como punto de interconexión con el módulo del cliente.

2) *Gestor del cliente de la comunicación:* La finalidad de este módulo es la de enviar la información capturada por el hooking al servidor, haciendo uso del módulo del protocolo. Se encarga de lanzar el hooking y controlar su correcta ejecución, dotándole de robustez frente a posibles fallos. A esto se añade que es el responsable de controlar la comunicación que se establece con el servidor y asegurar de que toda la información recibida es correctamente transmitida. Depende del módulo de hooking y del protocolo.

3) *Gestor del servidor de la comunicación:* Es la contraparte del anterior módulo y se encarga de recibir la información mandada por todos los clientes, a la vez que la clasifica para almacenarla en una base de datos. Está programado de forma que cada nueva conexión de un cliente esté asociada a una nueva hebra en el servidor. Depende del módulo del protocolo y de la base de datos.

4) *Protocolo de la comunicación:* Es el punto común a los módulos de cliente y servidor. Define un protocolo de mensajes y las funcionalidades necesarias para que pueda ser utilizado desde el nivel de aplicación, además de dotar de seguridad a la comunicación.

5) *Gestor de la base de datos:* Sirve como punto de acceso y gestión a una base de datos SQLite. Permite introducir toda la información recibida en dos tablas. La implementación permite el acceso de varias hebras, dando así soporte a las conexiones de varios clientes de forma simultánea.

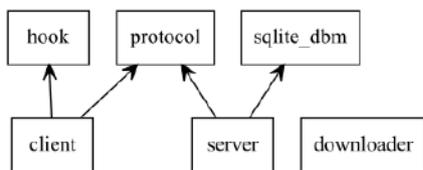


Fig. 1. División funcional del software desarrollado.

### B. Arquitectura de red: cliente-servidor

La elección de haber usado el esquema cliente-servidor ha sido tal porque responde fácilmente a la situación de varios sistemas infectados y un punto de reunión de la información.

Dado que este esquema empieza a mostrar problemas cuando la carga de clientes es muy grande, se puede optar por añadir más servidores a los que se gestiona el acceso con algún tipo de balanceo de carga. Estos nuevos servidores se comunicarían entre sí con otro protocolo para sincronizar la información recibida y evitar redundancias innecesarias.

Junto a la decisión del esquema de comunicación surgen otras cuestiones de diseño.

1) *Mutabilidad de la conexión:* Cuando se analiza el comportamiento de un posible de malware se presta especial atención a patrones que se repitan. En este caso, el hecho de que la comunicación se haga siempre con el mismo servidor con la misma dirección IP se presta a marcar más fácilmente cuál es la dirección relacionado con un comportamiento extraño, sino que además puede comprometer la estabilidad de los servidores mediante denegación de servicio o explotación. Para evitar esto, se puede optar por tomar las direcciones de los servidores de un diccionario integrado en el propio software o a partir de otras direcciones IP. En busca de una implementación simple y clara, no se han implementado medidas que protejan ante esta posibilidad.

2) *Protocolo de comunicación:* La búsqueda de un diseño simple pero válido lleva a elección de un protocolo propio formado por pocos paquetes pequeños, lo cual hace que sea más robusto lógicamente y minimiza su impacto en el sistema y en la red.

3) *Conexión con el servidor:* Los dos protocolos más relevantes en la capa de transporte son TCP y UDP. UDP es tradicionalmente asociado a conexiones rápidas donde se mandan paquetes de forma esporádica sin necesidad de esperar una respuesta, aunque no aplica ninguna medida de control de flujo ni detección de errores. El funcionamiento de UDP podría ser válido para el proyecto, pero dadas las facilidades de uso que aporta TCP se hace uso de este, anulando así la necesidad de comprobar metódicamente la correcta transmisión de los mensajes.

4) *Cantidad de conexiones:* Una vez que se ha elegido TCP como protocolo de transporte es necesario marcar cuál será la frecuencia de las comunicaciones, ya que TCP requiere de un *handshake* previo a la comunicación. La opción de utilizar una única conexión por instancia del cliente parece inicialmente la opción más simple, pero no por ello la mejor, pues esto afectaría al consumo de recursos en el servidor, además de denotar un comportamiento sospechoso en el que una conexión de gran duración manda pocos mensajes. Se elige por tanto un envío esporádico y se marca un tiempo entre conexiones para realizar una sola conexión en el caso de que se realicen varios accesos en poco tiempo.

5) *Ofuscación:* Aunque en el resto de apartados se ha tenido en cuenta la visibilidad del comportamiento del malware y cómo se puede ocultar en la medida de lo posible, aparecen algunas consideraciones adicionales que mejoran la ofuscación del comportamiento. Estas nuevas anotaciones son la definición del puerto al que se comunicarán los clientes, siendo este uno que pase por la mayoría de cortafuegos, como

es el puerto 80; y el cifrado de la conexión, que aunque idealmente debería hacer uso de protocolos como TLS, en términos de prueba de concepto se ha hecho uso de un cifrado AES.

### C. Protocolo

El protocolo a utilizar es parte importante de la comunicación, pues marca cuándo y cómo se intercambia la información. En un principio se podría haber hecho uso de uno ya creado o partir de un *framework* que facilite la implementación, aunque esto añadiría capas de abstracción que no son necesarias. Por ello se define un protocolo propio que responde a los requisitos pedidos de forma directa.

En primera instancia se optó por un protocolo que solo se encargase de servir la información obtenida por el hooking al servidor para que este la procesara. Pero esta dinámica dejaba atrás la posibilidad de compartir más información interesante que no fuese solo credenciales, sino valores relacionados con el sistema operativo del usuario y su navegador. Otra posible característica era la de marcar a cada usuario, definiendo así una entrada propio en la base de datos. Estas dos nuevas opciones se añadieron a la metodología inicial y se creó un protocolo que permitía dos modos de funcionamiento: uno tradicional y más simple que solo mandaba credenciales obtenidas por el hooking y otro que añadía a esa base, el envío previo de información del sistema y un reconocimiento de la existencia del usuario en la base de datos. A esta última opción se añadía la posibilidad de almacenar de forma local un identificador de usuario para que el envío de la información del sistema no supusiese un peor rendimiento.

## IV. IMPLEMENTACIÓN

Tras la etapa de diseño se ha realizado la implementación en Python teniendo en cuenta los parámetros decididos.

### A. Punto de hooking inicial

Antes de poder programar el módulo de hooking es necesario realizar un estudio del comportamiento del navegador y definir cuál es la dirección de memoria donde se realizará la toma de información.

La búsqueda de este punto de hooking o dirección de la función objetivo puede tomar varios caminos siendo todos ellos válidos. Para este proyecto se ha partido de la llamadas a funciones de red básicas implementadas por la librería `ws_32.dll`. Estas son: `send`, `recv`, `WSASend` y `WSARecv` para comunicaciones TCP y `sendto`, `recvfrom`, `WSASendTo` y `WSARecvFrom` para UDP. Dado que el protocolo a analizar es HTTPS, más concretamente el envío de información que se hace con él, se prestará atención a `send` y `WSASend`.

Una vez elegidas las funciones a estudiar, se realiza tráfico HTTPS con un navegador que tenga asociado un depurador. Se observa que para el caso de Firefox, de las dos funciones monitorizadas, solo `send` es utilizada, cuya sintaxis es la mostrada en el Listado 1.

```
int send(
    _In_ SOCKET s,
    _In_ const char *buf,
    _In_ int len,
    _In_ int flags
);
```

Listado de código 1. Sintaxis de la función `send` en C++

Como esta función es llamada numerosas veces, se toman solo las llamadas que contienen información clara de la petición HTTPS, aún sin cifrar, y se analiza su pila de llamadas. De esta forma se observa que todas las llamadas válidas a `send` cuentan con un llamada a `PR_Write` previa, cuya sintaxis es la mostrada en el Listado 2. Y la cual recibe un considerable menor cantidad de llamadas.

```
#include <prio.h>

PRInt32 PR_Write(
    PRFileDesc *fd,
    const void *buf,
    PRInt32 amount
);
```

Listado de código 2. Sintaxis de la función `PR_Write` en C++

Llegados a este punto se podría definir un punto de hooking correspondiente a la función `PR_Write` y cuyo segundo parámetro cuenta con la petición HTTP sin cifrar. Pero aunque esto es totalmente funcional, se da un paso más y se busca encontrar un punto de hooking con una mayor cantidad de positivos y una mejor lectura.

### B. Optimización de la llamada de hooking

Para conseguir la optimización de la llamada se vuelve a hacer uso de la pila de llamadas, pero esta vez centrando el foco en las llamadas que se estaban haciendo a `PR_Write`. De esta forma se obtiene que existen llamadas a funciones intermedias que se corresponden únicamente con envío de peticiones POST, aunque por contra, no tienen un nombre de función asignado y se ha de hacer referencia a ellas mediante su dirección de memoria.

Para resolver el tema de lecturas completas del buffer del mensaje sin perder información, se sigue el valor tomado por el tercer parámetro en la llamada a `PR_Write` y se identifica con el parámetro de igual valor en las nuevas funciones encontradas. Por último, solo queda obtener cuales de todas las posibles funciones intermedias son válidas, proceso el cual se realiza observando su comportamiento frente a varias páginas con HTTPS y tomando la que es común a todas ellas.

Conseguidos los valores de la dirección de memoria de la llamada y los parámetros a leer, se puede realizar la implementación con `PyDbg` en Python.

### C. Comunicación entre cliente y servidor

La implementación de la comunicación entre los extremos empieza como un esquema simple de cliente y servidor, pero al que se le van añadiendo más funcionalidades. Algunas de estas particularidades son:

- Se forma un flujo de decisiones en ambos lados para implementar el protocolo de mensajes especificado. Este

flujo es capaz de funcionar en los dos modos de ejecución diseñados sin necesidad de una configuración adicional.

- Se hace uso de `sendall()` para asegurar que llega todo el mensajes en el momento de decisión.
- Se añade un cierre de la comunicación con `shutdown(socket.SHUT_RDWR)` para notificar el cierre de la conexión al otro extremos y asegurar que se mande toda la información que aún no haya sido transmitida.
- Se tiene en cuenta que el servidor tendrá que dar servicio a varios clientes, por lo que en cada `accept()` del servidor se forma una nueva hebra que implementa el flujo de mensajes del lado del servidor.
- Se implementa el protocolo de mensajes teniendo en cuenta el cifrado de los mensajes. Dado que la implementación original era orientada a caracteres se hace una codificación en Base64 del mensaje cifrado.

#### D. Almacenamiento de la información

Para resolver el problema del almacenamiento de la información se hace uso de una base de datos SQLite gestionada únicamente desde Python. Dicha base de datos cuenta con dos tablas, una dedicada a entradas de usuarios con los parámetros recolectados de sus equipos y otra con las tuplas con formato `(site, user, password)` referentes a la credenciales usadas en cada página web. Ambas tablas pueden relacionarse entre sí mediante un identificador de usuario para el caso de que los usuarios hagan uso completo del protocolo. En el caso de que solo mandasen la información del hooking, solo se almacenarían sus entradas en la tabla de credenciales.

En cuanto al funcionamiento de la base de datos se ha tenido en cuenta que originalmente no se da soporte a acceso a esta desde varias hebras y que se ha de marcar de formar expresa que sí se pueda. Esta modificación no vienen sin coste añadido, y es que ahora se ha de programar de acuerdo a que el objeto que hace referencia a la base de datos sea tratado por solo una hebra a la vez, añadiéndole algo más de complejidad al código a cambio de la versatilidad obtenida.

La implementación final cuenta con métodos que permiten el ingreso de entradas en las tablas y la comprobación de la existencia previa de estas entradas.

#### E. Downloader

Se implementa un módulo lógicamente aislado del resto cuya funcionalidad es la de descargar y ocultar el verdadero malware en el sistema. La ocultación la consigue marcando el nombre del ejecutable del malware de forma similar a uno legítimo y localizándolo en una carpeta propia de la instalación del navegador. Para dotarlo de permanencia a lo largo de los futuros apagados, se añade una clave de auto-inicio al registro de Windows.

Este comportamiento es bastante notorio de cara a ser analizado por un antivirus, pero permite que se pueda usar un vector de infección distinto más apropiado para otros casos.

#### F. Empaquetado

Por último, se implementa el empaquetado de todo el proyecto y del Downloader. El empaquetado se realiza con

PyInstaller y se especifican como parámetros importantes que la salida sea en un solo archivo con nombre e icono personalizados y funcionamiento silencioso, es decir, que no lance una terminal ni ninguna interfaz cuando se ejecuta. Esto se aplica a los dos empaquetados, para la propagación solo restaría alojar el verdadero malware en algún sitio web y hacer uso del Downloader para que el malware se añada al sistema objetivo.

### V. EVALUACIÓN

El proyecto desarrollado ha conseguido ser la prueba de concepto que se buscaba, un proyecto que mostrase la viabilidad del uso de técnicas de hooking para la obtención de información confidencial. Se ha obtenido un producto que gracias a las consideraciones que se han hecho en la etapa de diseño consigue responder adecuadamente a situaciones de estabilidad y obtiene un funcionamiento silencioso propio del malware.

Aunque la implementación ha sido correcta, aun puede ser pulida para dar solución a problemática no tan propias de una prueba de concepto sino de un verdadero malware. Una de estas es que la implementación ha hecho uso de una dirección de memoria específica en lugar de hacer uso de funciones marcadas, esto conlleva que en futuras actualizaciones haya que revisar este proceso, que aunque mecánico, necesita ser rehecho. Para solucionarlo se podría optar por identificación de las funciones mediante huella o automatizar la búsqueda del punto óptimo. El haber optado por utilizar una comunicación propio es una solución altamente válida para un escenario de tamaño moderado, pero en casos de despliegue masivo se agradecería una implementación más completa. Y por último, el haber usado Python como lenguaje ha facilitado el proceso de desarrollo y permitido la inclusión de más opciones de forma más rápida a lo que se habría obtenido con un lenguaje más tradicional, pero por contra, al ser un lenguaje interpretado carga con la necesidad de requerir un intérprete, ya sea propio del sistema objetivo o auto-empaquetado, aumentado así la dependencia o tamaño del malware final.

En definitiva, el desarrollo es correcto para las dimensiones a las que se ha orientado, pero en el caso de que fuese a ser utilizado como un verdadero malware requeriría algunos ajustes.

Para completar el análisis se han escaneado tanto el propio el malware como el Downloader con varios antivirus comerciales. El malware ha sido detectado por 6 de 67 antivirus, siendo ninguno de los detectados usado comúnmente en entornos domésticos. Por otra parte, el Downloader ha sido detectado por 20 de 67, siendo este sí detectado por los antivirus domésticos, lo cual era esperable pues su implementación buscaba más demostrar las posibilidades que se podían conseguir sin entrar en el campo de la ocultación de su comportamiento.

### VI. CONCLUSIONES

Como conclusión de este proyecto se obtiene que las técnicas de hooking son válidas en la actualidad para el desarrollo de malware y que dado su carácter legítimo no son identificadas directamente como tal. Su desarrollo además es

más fácil en productos en los que se tiene acceso al código fuente, puesto que simplifican la etapa de reversing, y en el caso de los navegadores con mayor número de usuario, como lo son Firefox y Chrome, esto se cumple. Cuenta además con que la continua mejora de las prestaciones de los equipos hace que el impacto que debería de tener el detener la ejecución normal del código para ejecutar la función de hooking sea cada vez menos notorio, llegando a ser imperceptible por el usuario.

Aunque todo no podía ser bueno, pues el desarrollo de malware haciendo uso de estas técnicas requiere de un alto conocimiento del funcionamiento a bajo nivel de un ordenador, añadido al conocimiento necesario de reversing que se acrecienta si no se cuenta con el código fuente. Y aunque las prestaciones de los equipos actuales hacen que el funcionamiento de estas técnicas sea viable, no las deja sin la necesidad de obtener una implementación casi óptima y la definición de funciones de hooking que tengan en cuenta su impacto en la correcta ejecución del programa.

#### AGRADECIMIENTOS

Me gustaría agradecer a mi familia y amigos, aquellos que no han estado solo a mi lado durante la realización de este proyecto sino los que decidieron formar parte de mi vida y me han ayudado a crecer como persona, y con los que he podido contar cuando era necesario tomar un descanso de los estudios. Sin ellos muchas cosas no habrían sido posibles.

Junto a ellos, también es necesario agradecer a los profesores que a lo largo del grado han hecho posible que no decaigan mis ganas de interesarme en él y permitirme avanzar en áreas en las que no esperaba mucho apoyo.

Por último, pero no menos importante, a mi tutor, que comprendió desde el principio mi visión de trabajo y me ayudó a complementarla e hizo capaz que su desarrollo fuese asequible en todo momento.

#### REFERENCIAS

- [1] J. Seitz, "Gray Hat Python: Python programming for hackers and reverse engineers", No Starch Press, 2009.



# Optimización de un sistema de detección de amenazas en capturas de tráfico de red mediante algoritmos evolutivos

Tutor: Antonio Mora García, José Camacho Páez; e-mail: amorag@ugr.es, josecamacho@ugr.es

Titulación: Ingeniería Informática, Área de IT (Ingeniería Telemática)  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

Autor: Pablo Merino Ávila, e-mail: pablomerino25@correo.ugr.es

**Resumen**— Los ataques informáticos cada vez son más sofisticados y se realizan con mayor frecuencia. Esta realidad, se ha visto agravada exponencialmente con la situación que estamos viviendo actualmente a causa del COVID-19, cada vez, más usuarios se encuentran expuestos a las amenazas informáticas.

La solución más extendida de la ciberseguridad es la prevención, es decir, impedir la entrada de ataques en nuestra red. No obstante, en caso de que se hayan superado las medidas preventivas, será esencial detectar y responder a cualquier intrusión con eficiencia.

Este trabajo fin de grado, expone una técnica que se centrará en los sistemas de detección de intrusos basados en anomalías, en particular, en la Monitorización Estadística Multivariante de Red (MSNM) en su versión semi-supervisada.

Se aporta una optimización de MSNM basada en los Algoritmos Genéticos. Con ellos se pretende potenciar el aprendizaje semi-supervisado, calibrando el detector de ataques y consiguiendo mejorar su rendimiento.

**Palabras clave**— Análisis de Componentes Principales, Ciberseguridad, Sistema de Detección de Intrusos, Algoritmos Evolutivos, Detección de Anomalías.

## I. INTRODUCCIÓN

El empleo de sistemas de detección de intrusos conlleva la extracción de grandes cantidades de información, puesto que es necesario guardar todos los eventos que sucedan en la red. Esta cantidad masiva de datos supone una adversidad para poder extraer información útil.

Gracias al trabajo realizado por Camacho et. al en "PCA-based multivariate statistical network monitoring for anomaly detection" [3], se crea una técnica de detección de anomalías para redes informáticas, denominada Monitorización Estadística Multivariante de Red (MSNM, Multivariate Statistical Network Monitoring). MSNM se basa en el Control de Procesos Estadísticos Multivariante (MSPC, Multivariate Statistical Process Control). MSPC se fundamenta en el análisis de componentes principales y convierte los datos masivos de entrada en variables denominadas componentes principales.

Esta herramienta por sí sola presenta una carencia, y es que

precisa de una optimización para poder detectar los patrones que siguen los ciberataques.

Para ello, se propone en este proyecto la combinación de MSNM, con una optimización basada en los algoritmos genéticos, para así poder dar otra alternativa de optimización a las ya existentes y buscar mejorar su eficiencia computacional.

## II. ALGORITMOS GENÉTICOS

Los algoritmos Genéticos están basados en el funcionamiento de la selección y genética natural fomentando la supervivencia del más apto. Se clasifican dentro de los llamados algoritmos evolutivos [6].

Su fundamento trata de conseguir una solución a un problema partiendo de un conjunto de individuos, generado de manera aleatoria, denominado población. Cada uno de estos individuos equivale a una posible solución al problema, y van evolucionando según los principios de la selección natural.

De forma general, un algoritmo genético se basa en una población de soluciones codificadas, de forma similar a los cromosomas humanos. Estos cromosomas tendrán asociados un valor de bondad o fitness. En función de la dimensión de este valor, se le dará más o menos oportunidades de reproducción, por lo que se establece un proceso de selección del más apto.

Para el cálculo del fitness, se emplea una función denominada objetivo, (o función de evaluación), de modo que los organismos que mejor la resuelvan, obtendrán más probabilidad de ser seleccionados para reproducirse.

Un individuo o cromosoma es una combinación de valores o variables (genes).

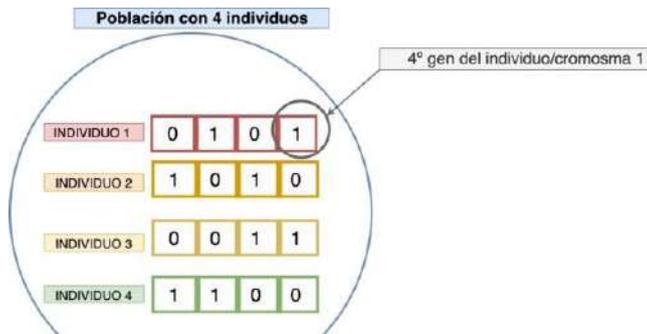


Fig. 1. Ejemplo de Población de Individuos. Fuente: elaboración propia.

La evolución se consigue a partir de la reproducción de los individuos y en esta pueden intervenir varios operadores que modifican la solución como pueden ser la selección, el cruce y la mutación.



Fig. 2. Fases del algoritmo genético. Fuente: elaboración propia.

#### A. Fase 0. Iniciar población.

Se crean una serie de individuos que serán el punto de partida del algoritmo, estos individuos formarán la población actual. Sus genes han de crearse con valores aleatorios.

#### B. Fase 1. Evaluación.

Esta fase se emplea para calcular la bondad de cada uno de los individuos de la población actual.

Esta fase es esencial para la selección, puesto que, al clasificar los individuos con un valor numérico, permite poder ordenar los individuos respecto a su fitness.

#### C. Fase 2. Selección.

En esta fase, los mecanismos de selección escogerán los individuos que van a obtener oportunidades de reproducirse, siendo los más competentes los que consigan mayor probabilidad de reproducción [1].

#### D. Fase 3. Cruce

Tras la selección se da paso a un proceso en el cual los individuos seleccionados se fusionan, para dar lugar a los nuevos integrantes que se agregarán a la nueva población. Los nuevos miembros de la siguiente generación, se denominan hijos. Los hijos heredarán (idealmente) algunas características de cada padre [9].

#### E. Fase 4. Mutación.

El proceso de mutación se basa en alterar uno o más genes de un individuo de la población. Esta alteración de cada gen se debe realizar dando lugar a cromosomas válidos.

#### F. Fase 5: Reemplazo

Esta estrategia, determina qué individuos estarán en la nueva población y a qué cromosomas sustituyen.

Para conseguir una evolución a lo largo de las generaciones, no se puede aumentar el tamaño de la población, ya que de lo contrario sería incontrolable la cantidad de información a gestionar. Hay que aplicar una

técnica que elimine algunos individuos de la población actual y los sustituya por los hijos que se han creado en los procesos previos de cruce y/o mutación [13], [14].

#### G. Condición de parada.

Es un estado el cual ha de establecerse previamente al inicio del proceso del algoritmo genético. Establece en qué situaciones terminará su trabajo iterativo.

### III. MSNM

Este proyecto se centra en la detección de anomalías, las cuales son eventos inusuales en el comportamiento de un sistema. Estos eventos pueden llegar a tener un efecto negativo dando lugar a errores, mal funcionamiento, o simplemente, una disminución del rendimiento en los dispositivos de interconexión [12]. Los dispositivos han de soportar una mayor carga de trabajo debido a las anomalías, pudiendo llegar a afectar a los clientes finales. Estas anomalías pueden ser provocadas por razones intencionadas o accidentales.

El Control Estadístico de Procesos (SPC, Statistical Process Control) es una metodología de detección de anomalías que originariamente empleaba estadísticas y gráficas con una única variable. Por tanto, sólo se monitorizaba una variable simultáneamente, lo que provocaba una pérdida de rendimiento del detector.

El problema del rendimiento dio lugar a emplear diversas variables simultáneamente, es decir, SPC Multivariante (MSPC, Multivariate Statistical Process Control). Este mecanismo, en un principio, no tenía en cuenta la relación latente entre las variables y esto llevó a aplicar métodos como el Análisis de Componentes Principales (PCA, Principal Component Analysis) [2] [3],[8].

PCA se aplica a conjuntos de datos con dos modos o dimensiones donde se calculan M variables o características para N observaciones. El objetivo de PCA consiste en encontrar el subespacio de máxima varianza en un espacio de M variables. El número de variables o Componentes Principales que se aplica se denomina PCs (Principal Components) y definen el rango del subespacio [3].

PCA es una herramienta que puede manejar un elevado número de variables o características de forma simultánea. De hecho, cuantas más variables existan, más información significativa extraerá sobre sus patrones [8],[4].

El análisis de componentes principales, sigue la siguiente expresión:

$$X := T_A \cdot P_A^t + E_A$$

Siendo "X" una matriz cuadrada de tamaño NxM elementos, se puede factorizar en unas nuevas variables, siendo estas las combinaciones lineales de la matriz original.

- A es el número de PCs. Cuantos más PCs se añadan, más parámetros tendrá el modelo.

- M es el número de variables diferentes tomadas en cada observación.

- N es el número de observaciones tomadas en total.

-  $T_A$  es la matriz de puntuaciones (scores), esta tiene un tamaño de NxM, y permite entender la estructura entre las observaciones.

-  $P_A$  es una matriz de carga (loadings), esta tiene un tamaño de MxA y permite visualizar la relación entre las variables.

Se representa en la expresión de PCA como  $P_A^t$ , donde la 't' denota la traspuesta de la matriz. Por tanto, el tamaño de P es  $A \times M$ .

-  $T_A \cdot P_A^t$  se denomina "el modelo". El producto de estas dos matrices,  $N \times A$  y  $A \times M$  respectivamente, tiene como resultado una matriz de tamaño  $N \times M$ .

-  $E_A$  es una matriz residual de  $N \times M$  elementos, donde se almacena lo que no capture el modelo.

La puntuación  $t_n$  en el modelo de una nueva observación  $x_n$  y se calcula con la siguiente expresión:  $t_n = x_n \cdot P_A$

Si se quiere calcular el residuo de dicha observación se debe usar la siguiente ecuación:  $e_n = x_n - t_n \cdot P_A^t$

La metodología MSPC basada en PCA se centra en los parámetros Q-estadístico (Qn), se calcula a partir de la matriz de residuos. Se representa en la *fig .3* (Q-st).

El D-estadístico (Dn), se obtiene a partir de la matriz de puntuaciones. Se representa en la *fig .3* (D-st).

Estos parámetros son complementarios y son calculados a partir de los datos de calibración, establecen así unos límites con unos Umbrales de Control (UCL, Upper Control Limit), haciendo posible clasificar como anomalías, los eventos que los superen en gran medida. Los UCL son las líneas discontinuas horizontales en la Figura *Fig. 3*. que se detalla más adelante.

Los estadísticos D y Q se pueden calcular para una observación n-ésima con las siguientes ecuaciones:

$$D_n = e_n \cdot e_n^t \quad Q_n = \sum_{m=1}^M (e_{nm})^2$$

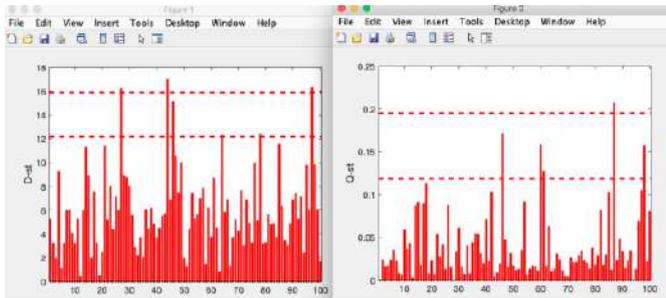


Fig. 3. Ejemplo D-estadístico y Q-estadístico. Fuente: elaboración propia, mediante la herramienta meda.

La Monitorización Estadística Multivariante de Red (MSNM, Multivariate Statistical Network Monitoring) consiste en un estudio realizado por José Camacho Páez [3]. Se trata de una técnica de detección de anomalías en redes informáticas, basándose en la técnica MSPC, pero para las redes informáticas. Alertará en caso de encontrar irregularidades en el tráfico de red, dando toda la información para identificar esa anomalía que se ha producido.

#### IV. CONJUNTO DE DATOS

El Dataset recogido en <https://nesg.ugr.es/nesg-ugr16/> es un conjunto de datos recopilados en la Universidad de Granada entre marzo y agosto de 2016. Este contiene trazas de tráfico de red junto con ataques informáticos modernos, es decir, capturas de todo el tráfico que se envía/recibe dentro de esa subred entre varios computadores.

El objetivo del dataset UGR'16 es evaluar herramientas de detección de ataques informáticos, concretamente IDS (Sistemas de detección de intrusos). Además, permite poner

en práctica diferentes algoritmos con grandes cantidades de datos, enfrentándose a casos concretos que tienen una extensión temporal realista.

Las capturas se dividen en dos:

- Capturas de Calibración: este primer conjunto, contiene datos comunes de tráfico y se emplean para entrenar. Este conjunto de datos se encuentra sin anomalías, por tanto, se han eliminado los valores atípicos (outliers) de las trazas en fase 1.
- Capturas de Test: contiene los datos de tráfico al igual que el primer conjunto, pero incluye algunos de los ciberataques más conocidos en la actualidad, emulados por investigadores del proyecto VERITAS. Permite poner a prueba la técnica de detección con un caso realista.

Los datos se obtienen de una infraestructura de red que tiene servicios virtualizados en la nube (Cloud). Esta nube es Trevenque, el mayor cloud de Andalucía. Aloja servidores de correo electrónico, de traspaso de ficheros por FTP (File Transfer Protocol) o incluso de Sistemas de Nombres de Dominio (DNS, Domain Name System). Este cloud alquila máquinas virtuales a sus clientes para que estos instalen todo tipo de servicios, por ejemplo, servidores web.

#### V. PROPUESTA CONCRETA

El presente proyecto combina el planteamiento no supervisado de MSNM, aprovechando su sensibilidad y mejorándolo mediante una optimización basada en algoritmos genéticos.

Con ello se brinda una opción más, que permite no solo reconocer amenazas, sino aprender los patrones que sigue una amenaza concreta. MSNM puede trabajar con grandes conjuntos de datos, obtenidos de redes de computadoras y se encarga de detectar anomalías en ellos. Se precisa de un conjunto de datos bien diseñado, el cual contenga tráfico real y ataques modernos. El conjunto de datos UGR'16 es ideal para esta propuesta, pues tiene numerosas ventajas, entre las que destaca su extensión temporal y su consideración a la naturaleza ciclo estacional de los datos del tráfico de red.

La información capturada en UGR'16, se almacena como vectores de características. Cada una de esas características es un contador del número de veces que ocurre un evento concreto en la red, en un espacio de tiempo determinado. Existen 134 características por cada vector en UGR'16.

El número de veces que ocurre cada evento por cada espacio de tiempo es muy dispar. Por tanto, se ha de dilatar o comprimir el número de muestras por cada unidad de tiempo determinada.

Todo esto origina una matriz de datos que tendrá tantas filas como observaciones se hayan realizado. Cada fila será un vector de características. Cada observación se encuentra etiquetada dependiendo de si en ese marco de tiempo, ha habido tráfico "normal", o si se ha producido un "ataque".

La optimización que se desarrolla en este trabajo se centra en la detección de cuatro tipos de ataques existentes en UGR'16. Esta analiza sus patrones de ataque y proporciona una técnica de aprendizaje, pudiéndose interpretar como una actualización del sistema basado en reglas.

Respecto a los cuatro ataques abordados, estos son los siguientes:

- Denegación de servicio (DOS)
- Escaneo de puertos (SCAN 11)
- Escaneo de puertos (SCAN 44)
- Tráfico de botnet (Nerisbotnet)

Por otra parte, para conseguir la optimización, se recurre a los algoritmos genéticos, donde se ha de partir de un conjunto de individuos o cromosomas, en el que cualquiera de ellos es una posible solución al problema. Para ello se lanza el algoritmo cuatro veces y se busca una solución para cada tipo de ataque descrito.

Cada individuo es una optimización de los parámetros de escala de MSNM. Los individuos están compuestos por genes, los cuales son una medida de la importancia relativa que tiene cada característica (peso). Por tanto, un individuo se puede definir como un vector de pesos.

A un conjunto de vectores de pesos se le llama población, por lo que habrá tantos vectores de pesos como número de individuos tenga la población. A su vez, un individuo se compone de 134 pesos diferentes, uno por cada característica.

Cada cromosoma de la población tiene asociado una bondad, fitness, o AUC, que equivale a la especificidad frente a la sensibilidad de detección y se expresa como un valor comprendido entre 0 y 1.

La bondad de un individuo se calcula mediante una función objetivo que se pretende maximizar con el algoritmo genético. Esta función permite clasificar los individuos de mayor a menor aptitud.

Los individuos evolucionan siguiendo las fases de los algoritmos genéticos, dando paso a nuevas poblaciones. Se establecen una serie de parámetros preestablecidos que afectan al desarrollo del proceso evolutivo, como son:

- Número de individuos: es un valor que indica cuántos cromosomas tiene la población actual.
- Número máximo de generaciones: determina cuándo termina el algoritmo, es decir establece la condición de parada.
- Tipo de ataque: es un valor que indica cuál de los 4 ataques se quiere detectar.

En cada iteración se desarrollan todas las fases del algoritmo genético.

Se crea una población inicial (fase 0) en la cual, cada uno de sus individuos se han generado con valores aleatorios comprendidos entre 0 y 1.

Tras ello, se evalúan los cromosomas de la población (fase 1) calculando su AUC. Al calcularlo, tendrá en cuenta cuál es el ataque concreto que se está buscando detectar.

Se ordenan los individuos de mayor a menor fitness, y se da paso a la fase de selección (fase 2). Esta fase determina cuáles son los individuos que van a tener oportunidad de reproducirse.

Para la selección se escogen la mitad de los individuos. Se experimenta con dos métodos de selección: inicialmente, escogiendo los individuos con mejor bondad y posteriormente, de forma aleatoria.

Este proceso continúa con una fase de cruce (fase 3). Se experimenta con dos tipos de cruce, el uniforme y el cruce en dos puntos. En cada uno de los cruces, se escogen dos cromosomas del conjunto de individuos seleccionados; estos generan dos nuevos hijos que se incorporan a la siguiente

generación. Una vez cruzados dos individuos no se pueden volver a cruzar más en esta generación.

Este valor está comprendido entre 0 y 1. Si vale 0, nunca se realizará el cruce y los hijos que se generen serán iguales a uno de sus dos padres. Por el contrario, si vale 1, la fase 3 (cruce) siempre se realizará.

Por cada cruce efectuado, se pasa a una fase de mutación (fase 4), en la que se alteran uno o varios genes en cada uno de los hijos generados. La mutación se realiza de acuerdo a una variable previamente definida que se denomina “probabilidad de mutación”.

El número de genes que se alteran, depende de una variable denominada “tasa de mutación”. El valor que se va a mutar se cambia por otro valor aleatorio entre 0 y 1.

Una vez generados los nuevos cromosomas que van a formar parte de la siguiente población, se ha de determinar, a qué individuos de la población actual sustituirán estos hijos. Es decir, se establece la fase de reemplazo (fase 5).

En esta fase 5 (reemplazo), el algoritmo sustituirá a la mitad de los individuos menos aptos, por los nuevos individuos generados.

Una vez creada la nueva población, se vuelve a la evaluación (fase 1), se realiza todo el proceso de evaluación de los individuos mediante la función objetivo. Ahora, la mitad de la población la constituyen los nuevos hijos, y la otra mitad, los padres de estos. A continuación, se procede a clasificar en función de su bondad, a los individuos que van a pasar a ser la población actual.

Se comprueba si se ha llegado a la condición de parada; en caso de que no se llegue, se volverán a repetir una vez más, todos los pasos del algoritmo genético (se pasará a la siguiente generación). En caso de que se cumpla dicha condición, se procede a establecer la solución final.

En caso de que la variable número de generaciones valga 0, únicamente se evaluará la población inicial y a continuación, se satisfará la condición de parada y se pasará a establecer la solución final.

El establecimiento de la solución final, consiste en un mecanismo que escoge el individuo más apto de la población actual.

Una vez establecida la solución de este ataque, se procede a buscar la solución para el siguiente tipo de ataque, es decir, se pasa de nuevo a la fase 0 (población inicial) pero, se cambia una variable que influye en la función objetivo, como es el “tipo de ataque”.

El algoritmo genético terminará cuando consiga establecer una solución para cada ataque concreto. Se obtienen así 4 soluciones, estas son cuatro individuos (o vectores de pesos) que pasarán a ser probados en un conjunto diferente de datos, junto con otras optimizaciones ya creadas.

Esta optimización se visualiza junto con el autoescalado de los datos (MSNM-AS) (El autoescalado es un tipo de preprocesado previo de los datos) y la propuesta de optimización de Camacho Páez et al. R2R-PLS.

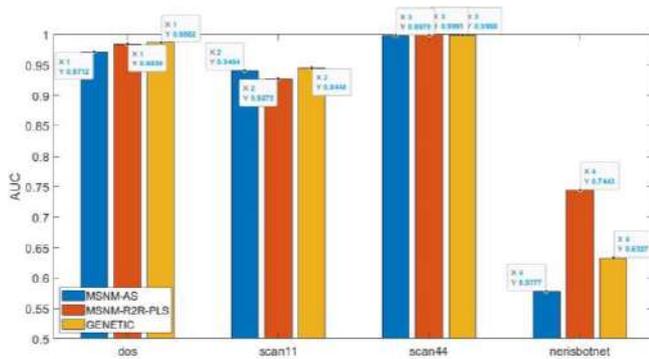


Fig. 4. Fitness en el conjunto de test de cada optimización. Fuente: elaboración propia.

En la fig. 4. se presenta el AUC conseguido para los 4 ataques de UGR'16, por tanto, figuran 3 tipos de optimizaciones:

- AS, (barra azul) denota autoescalado, un tipo de preprocesado que previamente se le ha realizado el centrado en la media. El autoescalado es aplicar MSNM sin optimizador.
- R2R-PLS (barra naranja) es la optimización basada en R2R-PLS.
- Genetic, (barra amarilla) es la optimización del algoritmo genético de esta propuesta.

Una vez terminado el algoritmo genético y comparado con todas las optimizaciones, se puede repetir el proceso tantos intentos como se quieran. El poder realizar varios intentos permite establecer una media de los resultados que se obtienen con el algoritmo.

Adicionalmente al algoritmo genético, se crearán una serie de metodologías para conocer factores que permitan interpretar, cómo se ha desarrollado el proceso de evolución, así como ver relaciones entre los individuos, conocer en qué generación se crea un individuo con mejor fitness y ser conscientes de la generación en la que se estanca.

Se ha creado una herramienta para exportar gráficamente la bondad del individuo mejor adaptado de la población actual (fitness máximo):

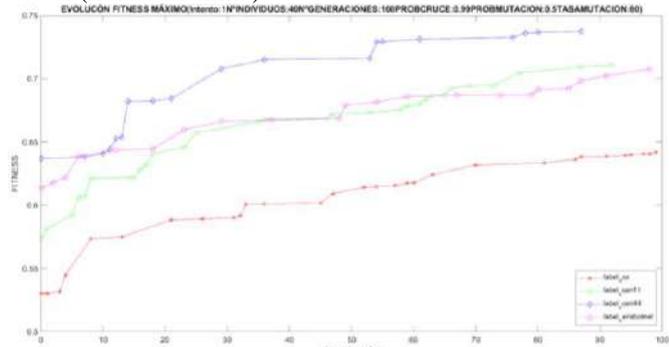


Fig. 5. Ejemplo de Gráfica de la evolución del fitness máximo de todos los ataques. Fuente: elaboración propia.

La Fig. 5. muestra en qué generación aparece un nuevo individuo con un fitness mayor que los vistos hasta el momento para los 4 tipos de ataque presentes en UGR'16. Detalla cuándo se ha creado un nuevo individuo en la población que da una mejor solución al problema.

Se ha implementado otra metodología para exportar gráficamente la bondad media de cada ataque, respecto al transcurso de las generaciones. La herramienta exporta gráficamente en total 4 gráficas de evolución del fitness medio. (Fig 6.)

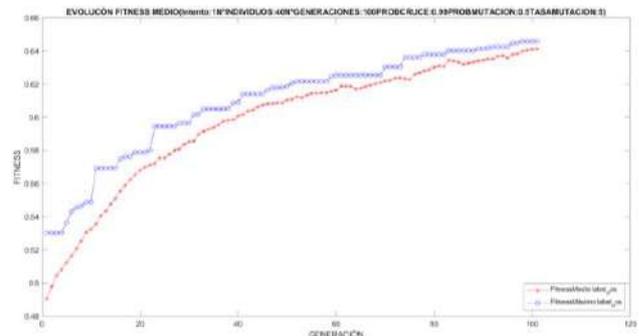


Fig. 6. Gráfica que muestra la evolución del fitness promedio de los individuos respecto al número de generaciones, junto con la evolución del fitness máximo respecto al número de generaciones, del ataque DOS.

Fuente: elaboración propia.

Por otra parte, se creará un método que exporte gráficamente la similitud entre los individuos de la población actual, respecto al número de generaciones (Fig. 7). La similitud es un valor comprendido entre 0 y 1. Si es 1, todos los genes de los individuos son iguales, si es 0 indica lo contrario.

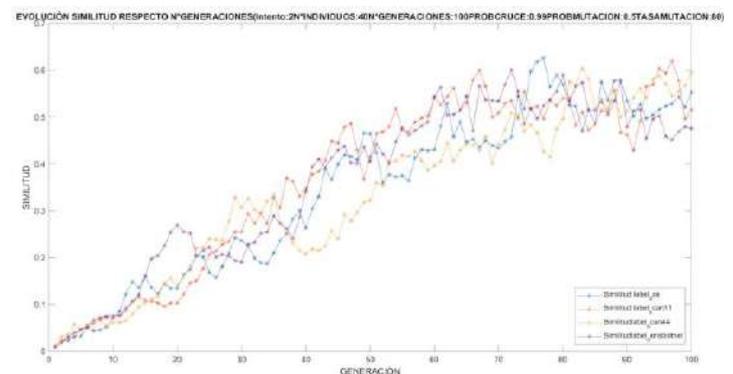


Fig. 7. Evolución de la similitud de la población actual respecto al número de generaciones transcurridas. Fuente: elaboración propia.

Se creará un método para exportar a hoja de cálculo los cromosomas, y la bondad que se ha obtenido de cada uno de ellos en el conjunto de calibración. Su objetivo es el de conocer mejor los genes de cada individuo observando conjuntos de genes iguales. Permite conocer por tanto el vector de pesos que se emplea como solución al algoritmo genético.

Cada vez que se ejecuta el algoritmo genético, se almacenan 4 hojas de cálculo, una para cada ataque. Se pone en la primera celda, el conjunto de características que han influido en la realización del algoritmo genético.

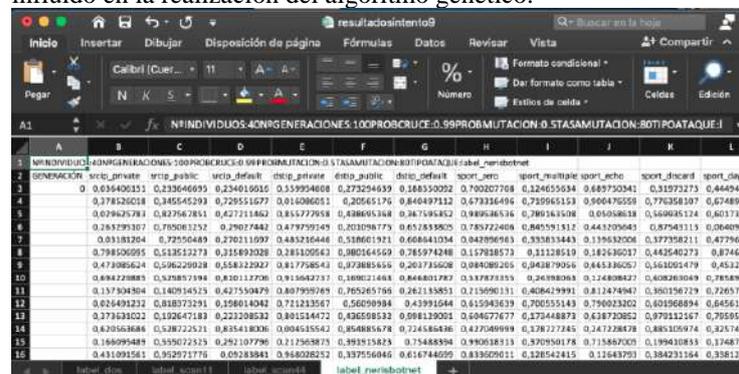


Fig. 8. Ejemplo de exportación a excel de unas poblaciones que optimizan MSNM para denegación de servicio (Nerisbotnet), Fuente: elaboración propia.

Dada la *fig. 8*, expone cómo a partir de la fila 3 el método para exportar, almacena los cromosomas de todas las generaciones, guardándolo en una hoja de cálculo diferente para cada ataque. La fila 2 indica la variable concreta que se optimiza en esa columna.

Los resultados obtenidos se almacenan automáticamente en una hoja de cálculo. Se guarda el fitness del mejor individuo hasta el momento, para cada ataque determinado (*fig. 9*).

Intento	As-dos	R2-dos	Gen-dos	As-scan11	R2-scan11	Gen-scan11	As-scan44	R2-scan44	Gen-scan44	As-bot	R2-bot	Gen-bot
1	0.9712	0.9849	0.9785	0.9404	0.9246	0.9439	0.9978	0.9986	0.9992	0.5777	0.7433	0.5343
2	0.9712	0.9831	0.9667	0.9404	0.9196	0.9179	0.9978	0.9976	0.9979	0.5777	0.7275	0.5213
3	0.9712	0.9855	0.9701	0.9404	0.9174	0.9244	0.9978	0.9992	0.9976	0.5777	0.7319	0.5547
4	0.9712	0.9853	0.9882	0.9404	0.8831	0.9411	0.9978	0.9892	0.9986	0.5777	0.7359	0.6083
5	0.9712	0.9657	0.9583	0.9404	0.9194	0.9334	0.9978	0.9892	0.9967	0.5777	0.7372	0.5461
6	0.9712	0.9807	0.9707	0.9404	0.9354	0.9359	0.9978	0.9974	0.9982	0.5777	0.706	0.5407
7	0.9712	0.9849	0.9808	0.9404	0.9168	0.9432	0.9978	0.9986	0.9985	0.5777	0.7447	0.688
8	0.9712	0.9869	0.9649	0.9404	0.9098	0.8853	0.9978	0.9993	0.9936	0.5777	0.7338	0.5383
9	0.9712	0.9681	0.9213	0.9404	0.8971	0.9315	0.9978	0.9986	0.9981	0.5777	0.7385	0.5728
10	0.9712	0.9833	0.9786	0.9404	0.9197	0.9227	0.9978	0.9976	0.9979	0.5777	0.7454	0.5375
media	0.9712	0.98461	0.967	0.9404	0.91419	0.92792	0.9978	0.99845	0.99757	0.5777	0.74042	0.55414

Fig. 9.: Ejemplo de exportación a excel de los fitness de los optimizadores, Fuente: elaboración propia.

En la *fig. 9*. se exponen, el mejor fitness obtenido en cada una de las 10 ejecuciones de las 3 configuraciones de optimización para MSNM: Autoescalado (AS), PLS-R2R (R2R), algoritmo genético (Gen).

Los resultados se almacenan por columnas denotando el ataque que optimizan: Denegación de servicio (dos), escaneo de puertos de una máquina a otra (scan11), escaneo de puertos de 4 máquinas a otras 4 (scan44), tráfico de botnet (bot).

## VI. CONCLUSIONES

Los algoritmos genéticos brindan una opción más que permite no sólo reconocer anomalías como lo hace MSNM, sino que también, logra aprender los patrones que siguen amenazas concretas. Todo ello supone una alternativa eficaz para optimizar el aprendizaje de MSNM.

El algoritmo genético consigue buenas optimizaciones para detectar los ataques de denegación de servicio (DOS) y los dos tipos de ataque de escaneo de puertos (SCAN11 y SCAN44) que se encuentran en el dataset UGR'16.

En concreto, el algoritmo genético generacional, es el modelo que aporta mejores resultados en la detección, frente al modelo estacionario que presenta los peores resultados.

Se han encontrado dificultades en algunos ataques como es en el tráfico de botnet, (Nerisbotnet) el cual se ha convertido en un problema deceptivo.

No se llega a conseguir los resultados obtenidos por Camacho Páez et. al con su optimización de R2R-PLS en cuanto a detección de la botnet, sin embargo, sí que iguala o incluso llega a superar en ocasiones dicha optimización en cuanto a la detección de los otros 3 ataques trabajados.

## REFERENCIAS

[1] ARRANZ DE LA PEÑA, Jorge; PARRA TRUYOL, Antonio. Algoritmos genéticos. Universidad Carlos III, 2007. Camacho, José and Pérez, Alejandro and Rodríguez, R. and Jiménez-Mañas, E. Multivariate Exploratory Data Analysis (MEDA) Toolbox, <https://github.com/josecamachop/MEDA-Toolbox>.

[2] Camacho, J., Macia-Fernandez, G., Fuentes-García, N. M., & Saccenti, E. (2019). Semi-supervised multivariate statistical network monitoring for learning security threats. *IEEE Transactions on Information Forensics and Security*, 14(8), 2179-2189.

[3] Camacho, J., Pérez-Villegas, A., García-Teodoro, P., & Maciá-Fernández, G. (2016). PCA-based multivariate statistical network monitoring for anomaly detection. *Computers & Security*, 59, 118-137.

[4] CAMACHO, Jose, et al. Tackling the big data 4 vs for anomaly detection. En 2014 IEEE conference on computer communications workshops (INFOCOM WKSHPs). IEEE, 2014. p. 500-505.

[5] Correa, C. A. P., & Díaz, H. P. (2007). Las amenazas informáticas: Peligro latente para las organizaciones actuales. *Revista Gti*, 6(16), 85-97

[6] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

[7] Gómez Hernández J. A. (2020). Diapositivas de seguridad en sistemas operativos, grado en ingeniería informática UGR.

[8] GURREA, Manuel. *Análisis de componentes principales*. Proyecto e-Math financiado por la Secretaría de Estado de Educación y Universidades (MECD), 2000.

[9] Herrera Triguero F. (2012). Diapositivas de bioinformática, grado en ingeniería informática UGR.

[10] MACIÁ-FERNÁNDEZ, Gabriel, et al. UGR '16: A new dataset for the evaluation of cyclostationarity-based network IDSs. *Computers & Security*, 2018, vol. 73, p. 411-424.

[11] RINGBERG, Haakon, et al. Sensitivity of PCA for traffic anomaly detection. En *Proceedings of the 2007 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*. 2007. p. 109-120.

[12] Lakhina, A., Crovella, M., & Diot, C. (2004). Diagnosing network-wide traffic anomalies. *ACM SIGCOMM computer communication review*, 34(4), 219-230.

[13] MARCOS; RIVERO GESTAL (DANIEL; RABUÑAL, JUAN RAMÓN; DORADO, JULIÁN; PAZOS, ALEJANDRO); GESTAL, Marcos. *Introducción a los algoritmos genéticos y la programación genética*. Universidade da Coruña, 2010.

[14] VALENCIA, Pablo Estévez. Optimización mediante algoritmos genéticos. En *Anales del Instituto de Ingenieros de Chile*. 1997. p. 83-92.



**Pablo Merino Ávila**, autor del presente trabajo, graduado en Ingeniería Informática en la universidad de Granada en 2020. Nacido en Granada el 25 de julio de 1997. Opositor al Cuerpo Nacional de Policía.



**Antonio M. Mora García**, tutor del presente trabajo, Doctor en Informática por la Universidad de Granada desde mayo de 2009, donde también obtuvo su Ingeniería Informática en 2001. En la actualidad trabaja como Profesor Titular de Universidad dentro del Departamento de Teoría de la Señal, Telemática y Comunicaciones en la misma universidad.



**José Camacho Páez**, cotutor del presente trabajo, Doctor en Informática por la Universidad de Granada desde 2007, donde también obtuvo su Ingeniería Informática en 2003. En la actualidad trabaja como Profesor Titular de Universidad dentro del Departamento de Teoría de la Señal, Telemática y Comunicaciones en la misma universidad.

# Composición de Cadenas de Servicios con un Algoritmo de Optimización Basada en Colonias de Hormigas

Tutor: Antonio Miguel Mora García; e-mail: amorag@ugr.es

Titulación: Doctor en Ingeniería Informática

**Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada**

Autor: Segundo Moreno Torres, e-mail: segundomoto@correo.ugr.es

**Resumen**— Debido al gran crecimiento de tráfico de datos a la par que la demanda de servicios de nueva generación se plantean nuevos retos a tener en cuenta a la hora de diseñar las redes de próxima generación. Una técnica que permitirá la ejecución de servicios avanzados, siendo el tráfico de red dirigido gracias a una lista de funciones ordenada será SFC (*Service Function Chaining*). Su importancia está en auge dado el crecimiento de SDN (*Software Defined Networks*) y NFV (*Network Function Virtualization*), junto al gran abanico de posibilidades que ofrecen en cuanto respecta a flexibilidad y automatización. En este TFG se plantea el desarrollo del diseño e implementación de un algoritmo de Optimización basado en Colonias de Hormigas (OCH) para la minimización del coste de routing de las cadenas de servicios, gracias a esta metaheurística que está especialmente diseñada para trabajar con grafos con pesos.

**Palabras clave**—ACO, Metaheurísticas, NFV, OCH, SDN, SFC

## I. INTRODUCCIÓN

Dado el cambio continuo que están sufriendo las redes de comunicaciones en estos últimos años, está cambiando completamente la forma en la que accedemos a Internet, bien sea a través de smartphones o cualquier dispositivo inteligente. Nuevas aplicaciones y servicios están surgiendo para aportar nuevas funcionalidades, como por ejemplo multimedia UHD (*Ultra High Definition*) o la opción de juegos *on-line*. Todo ello requiere de una gran velocidad en la red además de una baja latencia, lo cual podrán proporcionar las redes de quinta generación (5G).

Para dicho fin, deberán desarrollarse nuevas tecnologías a nivel físico que puedan soportar mayores tasas de tráfico de datos, despliegues de UDN (*Ultra-Dense Networks*) y un uso de frecuencias más altas para los mismos. A un nivel superior mirando en la pila de protocolos, actuarán SDN y NFV, donde gracias a estas dos metodologías podrán implementarse redes que eviten el uso de hardware específico para ofrecer determinados servicios, dando así mayor flexibilidad a las mismas a la par de una gestión mucho más clara y eficiente.

Los servicios de dichas redes serán ofrecidos gracias a SFC (*Service Function Chaining*), o cadenas ordenadas de funciones, siendo estas formadas por un conjunto de módulos virtualizados de software para poder ejecutarse en distintos nodos de la red [1].

El desarrollo de este trabajo se ha enfocado en SFC, planteando la composición de dichas cadenas de servicios como un problema de tipo NP-Completo [2]. El problema

se definirá a partir de un grafo que modela la red, donde se despliegan nodos capaces de ofrecer distintas funciones cada uno. Cada enlace entre estos nodos tendrá una capacidad limitada de ancho de banda. Se recibirán distintas peticiones de conexión para satisfacer a diversos usuarios, que irán siendo procesadas para la composición de una cadena siguiendo un orden determinado. Así, este problema se convierte en uno de búsqueda de camino óptimo dentro de la red, intentando encontrar la secuencia de funciones deseada utilizando el mínimo de recursos posible y dando el menor número de saltos entre nodos. Cada conexión corresponderá a un problema de este tipo, que deberá tener en cuenta las conexiones anteriores y los recursos de la red actuales.

Para la resolución de este problema se aplicará OCH, una metaheurística que proporciona gran efectividad en la búsqueda de camino óptimo en grafos [3][4][5].

Es por ello que en el trabajo se ha desarrollado una variante de OCH adaptado para la resolución de este problema con las restricciones requeridas. Teniendo esto en cuenta, se ha denominado *Ant-SFC*. Este ha sido validado en tres escenarios, uno a modo de prueba de concepto y otros dos modelando esquemas más realistas.

## II. PROBLEMA: ROUTING PARA SFC

NFV ofrece nuevos mecanismos para el diseño y planificación de redes gracias a la virtualización de servicios. Sin embargo, estas ventajas vienen acompañadas de algunos retos a resolver, como pueden ser el desarrollo de técnicas para la ubicación de manera dinámica de las funciones virtuales [6] además del desarrollo de algoritmos para la optimización del routing dentro de la red [7].

En este trabajo nos centraremos básicamente en el segundo de estos retos indicados, relacionado con la construcción de una cadena de servicio (SFC) dentro de la red. Es por ello que este problema ha sido denominado *Optimización del Routing en SFC* (OR-SFC).

Cada conexión vendrá definida por una tupla donde se incluirán los valores de origen, destino, valor de la demanda y las funciones virtuales a ejecutarse. El nodo de origen y destino serán los correspondientes a donde comenzará y terminará la conexión, mientras que el valor de demanda equivaldrá a las exigencias de recursos (CPU, memoria, etc.) que deberá tener

dicho nodo y su enlace disponible. Las funciones virtuales deberán servirse en el orden indicado. Es por ello que para calcular el camino deberán cumplirse estas condiciones, además de otras restricciones como que nodos y enlaces dispongan de suficientes recursos y capacidad respectivamente para soportar la demanda.

### III. ESTADO DEL ARTE

Para la formación de cadenas SFC una de las mayores complicaciones es la propia construcción del camino a seguir por el tráfico. Este problema, planteado como uno de tipo NP-Completo [2] ha llamado la atención por parte de la academia, siendo posible clasificar las soluciones que lo abordan en dos tipos: soluciones exactas o métodos heurísticos.

Los métodos que aplican soluciones exactas se centran en modelos de optimización basados en técnicas de programación lineal en búsqueda de soluciones exactas. A pesar de que ofrecen soluciones precisas, estas técnicas no son aplicables en instancias grandes del problema, debido a la gran cantidad de recursos de cálculo y tiempo que necesitarían.

La otra manera de resolver este tipo de problemas es mediante soluciones heurísticas (o metaheurísticas). Con ellas es posible obtener soluciones muy próximas a los óptimos en tiempos de cómputo muy reducidos. A pesar de que son poco flexibles y ante cambios dentro de la instancia no se podrían adaptar de manera sencilla, las técnicas que usan metaheurísticas son más adecuadas para trabajar con instancias de gran tamaño, donde no es posible métodos que trabajen con soluciones exactas.

Este trabajo es el primero propuesto que ofrece una solución basada en Colonia de Hormigas para la optimización de routing en SFC, pretendiendo así ser una primera propuesta para que en un futuro sea posible desarrollar soluciones más completas y avanzadas.

### IV. OPTIMIZACIÓN BASADA EN COLONIAS DE HORMIGAS (OCH)

OCH u Optimización Basada en Colonias de Hormigas es una metaheurística que se inspira en el comportamiento de las hormigas para formar los caminos más cortos posibles para llegar desde su nido hasta una fuente de alimento [8], guiadas por la *estigmergia*, equivalente a la comunicación entre agentes a través del medio en dónde se sitúan. Esto sucede gracias al depósito de *feromona* que cada hormiga va realizando conforme va formando un camino. El resto de las hormigas podrán seguir este rastro y en función de la intensidad de dicho rastro, más hormigas tenderán a seguir el camino formado. Esta feromona se evapora con el tiempo, con lo que es necesario una fuerte realimentación (aporte) de feromona para que el camino mantenga su rastro. Por ello, el camino que más se refuerza será el más corto, ya que más hormigas tenderán a pasar por este.

Pasando del mundo natural al artificial, la colonia de *hormigas artificiales* se comunica a través de una *matriz de feromona*. En cada iteración, cada una de las hormigas construirá un camino completo (una solución) y depositará feromona a lo largo de este. Dicho rastro será una medida informativa para el resto de hormigas, donde se les indica cuán bueno será el camino creado.

Cada hormiga tendrá disponibles dos tipos de información para moverse por el grafo: la *información memorística* y la *información heurística*. La primera de estas corresponde a la feromona depositada, mientras que la segunda dependerá del problema y se basa en el conocimiento previo del mismo.

Las hormigas, en su movimiento a lo largo del grado, irán eligiendo los nodos con mejor valor para dichas informaciones, aunque existirá una componente estocástica por el que una hormiga tendrá opción de moverse a un nodo de menor valor, favoreciendo así la exploración.

Para este trabajo ha sido utilizado un desarrollo básico de OCH, el *Ant System* o Sistema de Hormigas (SH). La estructura puede verse en los pseudocódigos de Algoritmo 1 y Algoritmo 2

#### Algoritmo 1: Sistema de Hormigas (SH) ()

Algoritmo principal de un SH

```
Inicializar parametros ()
while criterio de terminacion no satisfecho
do
  for cada hormiga h do
    s[h]=Construir Solucion(h)
  end for
  /* En todos los arcos que forman el grafo
  */
  Evaporacion de la Feromona ()
  /* Sólo en los arcos del mejor camino
  (mejor hormiga */
  s*=Elegir Mejor Solucion(s[h])
  Actualizacion Global de la Feromona(s*)
end while
```

#### Algoritmo 2: Construir Solución (id hormiga)

Algoritmo para la construcción de una solución por parte de una hormiga en un SH genérico

```
Inicializar hormiga(id hormiga)
Estado actual = estado inicial
L = guardar(estado inicial) /* Lista de
estados visitados */
while estado actual ≠ estado objetivo do
/* A: lista de estados alcanzables, P:
probabilidad de cada estado alcanzable, Ω:
restricciones del problema */
P = calcular probabilidades de
transicion(estado actual, A, L, Ω)
estado siguiente = aplicar politica decision
(P, Ω)
mover al siguiente estado (estado siguiente)
L = guardar (estado siguiente)
estado actual = estado siguiente
end while
```

Como es posible apreciar, el algoritmo basa su funcionamiento en una repetición continua de un proceso sencillo, hasta alcanzar el número indicado de iteraciones. Cuando todas las hormigas terminan de construir su camino, se procederá a actualizar la feromona de los arcos que forman el grafo. Esto se realizará mediante un proceso de evaporación, aplicado en todas las hormigas y caminos, además de un aporte final a la mejor hormiga y su solución, reforzando así los arcos más ventajosos.

### V. ALGORITMO GENERADO: ANT- SFC

Como se ha comentado anteriormente, la solución ha sido la implementación de un Sistema de Hormigas, aunque además se le han añadido una serie de restricciones necesarias para solventar nuestro problema:

- El camino que se defina dentro del grafo deberá pasar por los nodos que sean capaces de servir las funciones de red que se requieren justo en el orden indicado.
- Los enlaces por los que pasen los caminos deberán tener un ancho de banda disponible suficiente para poder responder a la demanda de tráfico que viene asociada con cada conexión solicitada. Este ancho de banda irá disminuyendo conforme más conexiones (y sus caminos) pasen por cada enlace.
- Los nodos en los que se sirvan las funciones deberán tener recursos suficientes en cada momento. Estos recursos irán disminuyendo a la par que más funciones de red sean ejecutadas en estos nodos.

El planteamiento del problema se ha hecho a modo de un problema de búsqueda de camino óptimo por separado, aunque realmente sean dependientes dado el consumo tanto de ancho de banda en los enlaces como de recursos en los nodos mencionado anteriormente.

El modelo principal de *Ant-SFC* quedaría tal cual se ve en Algoritmo 3, mientras que el pseudocódigo equivalente a la construcción de una solución se puede apreciar en Algoritmo 4. La modificación y adaptación del algoritmo se ha centrado en los siguientes aspectos:

- *Inicialización de la red*: tanto anchos de banda y recursos de la red deberán estar inicializados al estado que tenía antes de que la hormiga anterior la modificase al construir su camino.
- *Heurística*: en el sentido estricto de la palabra no se ha considerado una heurística. Sin embargo, sí que se ha asignado una mayor probabilidad de ser elegidos a aquellos enlaces con un mayor ancho de banda disponible (reduciendo el riesgo de agotamiento de ancho de banda). También se ha añadido una condición que hará que, si un nodo tiene disponible la siguiente función a servir, su probabilidad de ser elegido será el doble.

#### - **Algoritmo 3: Ant-SFC ()**

*Algoritmo principal de Ant-SFC*

```
Inicializar parametros ()
Leer configuracion red ()
Leer conexiones()
/*Se busca una solución por cada conexión */
for cada conexion c do
  while criterio de terminacion no satisfecho
  do
    for cada hormiga h do
      s[h]=construir.solucion (c, h)
    end for
    /* En todos los enlaces del grafo */
    Evaporacion de Feromona()
    /* Enlaces recorridos por la mejor hormiga */
    s*=Elegir Mejor Solucion(s[h])
    Actualizacion Global de Feromona(s*)
  end while
  /* Se actualizan anchos de banda en los
  enlaces y recursos disponibles en los nodos
  */
  Actualizar Red (c, s*)
end for
```

- *Nodos alcanzables*: la lista de los nodos para los que existe posibilidad de alcanzarlos desde un enlace actual se ha definido como 'A', teniendo en cuenta que dispondrá

de suficiente ancho de banda y que el nodo seleccionado dispondrá de la siguiente función virtual a servir por la cadena de funciones.

- *Ruleta de probabilidad*: cuando es asignada la probabilidad de moverse a un nodo desde el actual, se usará una ruleta para decidir el estado siguiente, basándose su funcionamiento en la asignación de probabilidad de cada nodo en una 'ruleta virtual'.
- *Restricción de ancho de banda en un enlace*: necesaria para la construcción de la lista de nodos alcanzables.
- *Restricción de recursos en el nodo*: al igual que en el caso anterior, es una condición necesaria para la construcción de la lista de nodos alcanzables.
- *Actualización de enlaces y nodos (construcción de la solución)*: cada vez que una hormiga se mueva a un nodo dentro de la red mientras está formando su camino para la resolución de una conexión, se actualizará el ancho de banda de dicho enlace y los recursos restantes de los que dispondrá el nodo (si es que este sirve alguna función), evitando así la formación de bucles infinitos dentro de la construcción de una solución.
- *Actualización de la red (conexiones)*: cada vez que se encuentre una solución para una cadena, toda la red se actualizará teniendo en cuenta el camino calculado en dicha solución (enlaces y nodos atravesados).
- *Restricción de camino completo*: una solución sólo se considerará como válida si comienza y termina en los nodos indicados en la tupla de conexión procesada, además de que haya servido todas las funciones virtuales en el orden requerido. En caso contrario, no se considerará dicha solución como válida y será eliminada.

#### **Algoritmo 4: Construir Solución (conex, id hormiga)**

*Algoritmo de construcción de una solución en Ant-SFC*

```
inicializar hormiga (id hormiga)
inicializar red () /* Fijar valores actuales de
la red */
nodo actual = conex.nodo inicial
funcion actual = conex.funciones[inicio]
L = guardar (nodo actual)
/* Lista de estados visitados */
F = guardar (funcion actual)
/* Lista de funciones servidas */
while (nodo actual ≠ conex(nodo final)) AND
(funcion actual ≠ conex.funciones[final]) do
/* A: lista de nodos alcanzables, P:
probabilidad de moverse a cada nodo alcanzable,
Q: restricciones del
problema */
P = calcular probabilidades de transicion (nodo
actual, A, F, L, Q)
nodo siguiente = ruleta probabilidad (P, Q)
/* Se actualizan anchos de banda en los enlaces
*/
Actualizar Enlace (nodo siguiente)
L = guardar (nodo siguiente)
nodo actual = nodo siguiente
/* Si está la función se sirve y se actualizan
los recursos en el nodo */
if funcion actual in nodo actual.funciones[]
then
  Actualizar Nodo (funcion actual)
  F = guardar (funcion actual)
  funcion.actual =
conex.siguiente(funciones[])
end if
end while
```

- *Coste de un camino (conexión):* el coste a tener en cuenta será equivalente al número de saltos requeridos dentro del grafo para completar la cadena de funciones requerida en cada caso.
- *Coste global de una solución:* una solución completa estará formada por el conjunto de caminos de mínimo coste procesados para resolver las conexiones requeridas. Es por ello que será la suma del coste de los caminos de todas las conexiones que se tengan.
- *Actualización de feromona:* este aporte será usado únicamente en los enlaces de la mejor solución procesada (menor coste), siendo este proporcional al número de saltos. Se usa junto con la evaporación de feromona, que se aplica a todos los enlaces del grafo. Se podrá deducir que, a menor coste total, mayor aporte realizado en los enlaces del camino.

VI. EXPERIMENTOS Y RESULTADOS

A continuación, se describirán las tres instancias procesadas en este trabajo y que han sido abordadas por el algoritmo.

A. Descripción de las instancias:

Se han evaluado tres instancias distintas del problema:

- Instancia 6N: grafo con 6 nodos, utilizado a modo de prueba de concepto y donde se ha podido evaluar el algoritmo de un modo más sencillo. En la Ilustración 1 se puede observar cómo está planteado.
- Instancia 19N: grafo con 19 nodos, modela un escenario más cercano a un real, más complejo que el anterior. Es posible observar su topología en la Ilustración 2, así como los datos asociados a los nodos y enlaces en la Ilustración 4.
- Instancia 52N: grafo con 52 nodos, equivalente a un escenario real de una red 5G, mucho más complejo de analizar y procesar pero que arroja los resultados más realistas. Para este caso se adjuntan los resultados en la Tabla 4. La configuración inicial utilizada se puede apreciar en la Tabla 1.

Parámetro	Instancia 6N	Instancia 19N	Instancia 52N
Num. Iteraciones	6	19	52
Num. Hormigas	12	38	104
$\alpha$ (peso feromona)	1.2	1.2	1.2
$\beta$ (peso heurística)	2.0	2.0	2.0
$\rho$ (factor evaporación)	0.3	0.3	0.3

Tabla 1: parámetros considerados en los experimentos

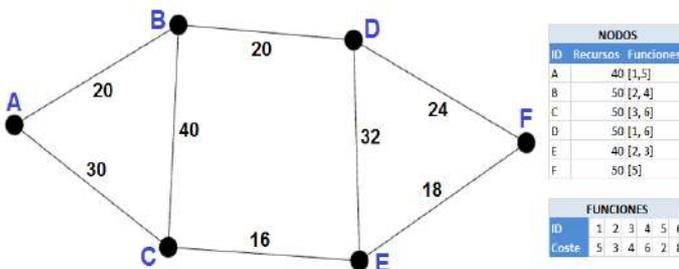


Ilustración 1: instancia de 6 nodos. Se puede apreciar la topología de red, nodos y ancho de banda del que disponen. Se aprecian nodos con funciones asociadas y recursos disponibles, además de los costes de cada función virtual.

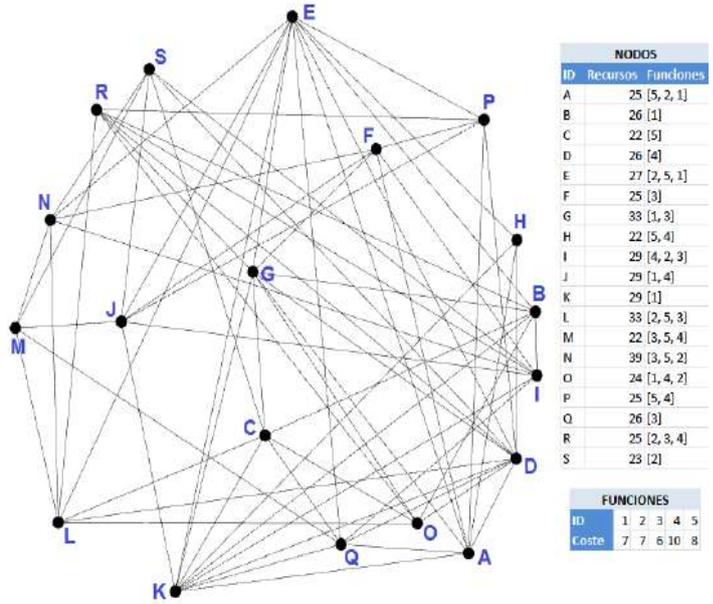


Ilustración 2: Instancia de 19 nodos: Topología de la red. Nodos con funciones asociadas y los recursos totales. Funciones virtuales con su coste asociado

B. Resultados obtenidos:

En la realización de este trabajo se ha utilizado una máquina con procesador Intel Core i7-4510 U de 2 núcleos y 4 hilos a 2.00 GHz, 8GB de memoria RAM DDR-3 con un S.O. Windows 10 de 64 bits.

Los resultados pueden observarse en las tablas aportadas en este trabajo. La primera de estas (Tabla 2) es la correspondiente a la ejecución de la instancia de 6 nodos, donde la mejor solución obtenida requiere de 13 saltos.

Para la primera instancia (6N) se tendrán que resolver las siguientes tres conexiones:

- Conexión 1: (A, F, 2, [3,5,6])
- Conexión 2: (A, E, 8, [1,2,4])
- Conexión 3: (A, D, 5, [2,4,5])

Dado que se trata de un algoritmo no determinista, se han realizado 10 ejecuciones completas de cada instancia para comprobar su correcto funcionamiento (con las mismas conexiones en cada caso).

En la Ilustración 3 se pueden observar los caminos seguidos dentro de la instancia de 6N para llegar a formar una solución completa que sirva las 3 conexiones exigidas y con todos los requisitos necesarios. Se pueden comprobar los resultados en la Tabla 2, donde se observa que la mejor solución requiere de 13 saltos y se da en la cuarta ejecución.

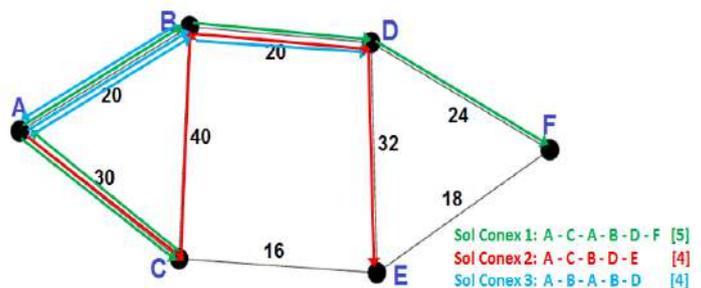


Ilustración 3: Mejor Solución encontrada para la instancia de 6 nodos con 3 conexiones. Coste en número de saltos junto a cada conexión.

Nº de ejecución:	Conexión 1	Conexión 2	Conexión 3	Solución global 6 nodos
1	8	3	5	16
2	12	4	4	20
3	11	4	4	19
4	5	4	4	13
5	7	4	4	15
6	8	5	5	18
7	6	3	5	14
8	7	3	4	14
9	12	4	6	22
10	20	3	4	27

Min. coste solución:	13
Peor coste solución:	27
Desv. Típica de todas las soluciones:	4,37
Media de todas las soluciones:	17,8

Tabla 2: datos de las simulaciones para la instancia de 6N y resultados.

En lo que respecta a la segunda instancia (19N) se tendrán que resolver las siguientes conexiones (en este caso 5 conexiones):

- Conexión 1: (H, J, 8, [5,1,2])
- Conexión 2: (B, D, 8, [4,3,1])
- Conexión 3: (Q, B, 1, [2,3,1])
- Conexión 4: (R, J, 3, [5,2,3])
- Conexión 5: (J, S, 8, [4,1,3])

Tras realizar de nuevo 10 ejecuciones, se pueden observar los resultados en la Tabla 3 pudiendo apreciar que el mejor resultado se obtiene con 20 saltos. En la Ilustración 5 se pueden observar los caminos seguidos en el cálculo de la solución global para esta instancia, apreciándose 5 caminos ya que se le exigen 5 conexiones distintas. En este caso se tiene una desviación típica mucho menor (2,33), considerando que se obtienen mejores soluciones incluso que en la primera instancia (6N) realizada.

Para la tercera y última de las instancias (52N) los resultados obtenidos se ofrecen en la Tabla 4, observando el mejor de los resultados en la ejecución número 9 con un coste total de 28 saltos. En este caso no se ha generado un gráfico ya que sería de poca ayuda dado el tamaño del grafo resultante, aunque cabe decir que es relevante a nivel de resultados, dado que para un problema más complejo se obtienen resultados muy buenos, observando además una desviación típica de sólo 1,77.

De especial relevancia son los tiempos medios de ejecución para todas las instancias. Para la primera de ellas, el tiempo de ejecución ha sido de 0,138s, para la segunda de estas de 0,389s mientras que para la última se ha obtenido un resultado de 5.669s. Si bien los resultados obtenidos de la primera y segunda instancias podrían considerarse como potenciales para ejecuciones en tiempo real y a pesar de que el resultado de la tercera instancia no lo parezca, se podrían establecer limitaciones de tiempo en función de las necesidades requeridas. Es importante recordar que en cualquiera de las instancias se ofrecen soluciones perfectamente válidas desde el primer camino de la primera iteración ejecutada.

ENLACES				ENLACES			
Nodo Orig	Nodo Dest	Peso	Ancho Banda	Nodo Orig	Nodo Dest	Peso	Ancho Banda
A	D	1.0	186	E	L	1.0	195
A	E	1.0	105	E	N	1.0	194
A	F	1.0	113	E	P	1.0	176
A	G	1.0	146	E	Q	1.0	120
A	H	1.0	185	F	G	1.0	164
A	K	1.0	116	F	I	1.0	189
A	P	1.0	101	F	J	1.0	136
A	Q	1.0	137	F	N	1.0	136
B	C	1.0	143	F	P	1.0	125
B	E	1.0	190	G	K	1.0	150
B	G	1.0	141	G	O	1.0	114
B	I	1.0	173	G	R	1.0	172
B	K	1.0	142	H	K	1.0	194
B	O	1.0	147	I	J	1.0	193
B	R	1.0	113	I	K	1.0	108
C	G	1.0	132	I	N	1.0	112
C	K	1.0	164	I	Q	1.0	155
C	L	1.0	184	I	R	1.0	162
C	O	1.0	118	I	S	1.0	104
C	Q	1.0	178	J	K	1.0	100
C	S	1.0	145	J	M	1.0	110
D	E	1.0	180	J	P	1.0	162
D	G	1.0	123	J	S	1.0	111
D	H	1.0	122	K	Q	1.0	141
D	K	1.0	139	L	M	1.0	160
D	L	1.0	127	L	N	1.0	176
D	O	1.0	165	L	O	1.0	135
D	P	1.0	179	L	R	1.0	196
D	Q	1.0	111	M	N	1.0	115
D	R	1.0	137	M	Q	1.0	159
D	S	1.0	188	M	S	1.0	161
E	G	1.0	160	N	S	1.0	166
E	H	1.0	200	O	R	1.0	191
E	J	1.0	186	P	R	1.0	124
E	K	1.0	150				

Ilustración 4: Instancia de 19 nodos: Anchos de banda asociados a los enlaces.

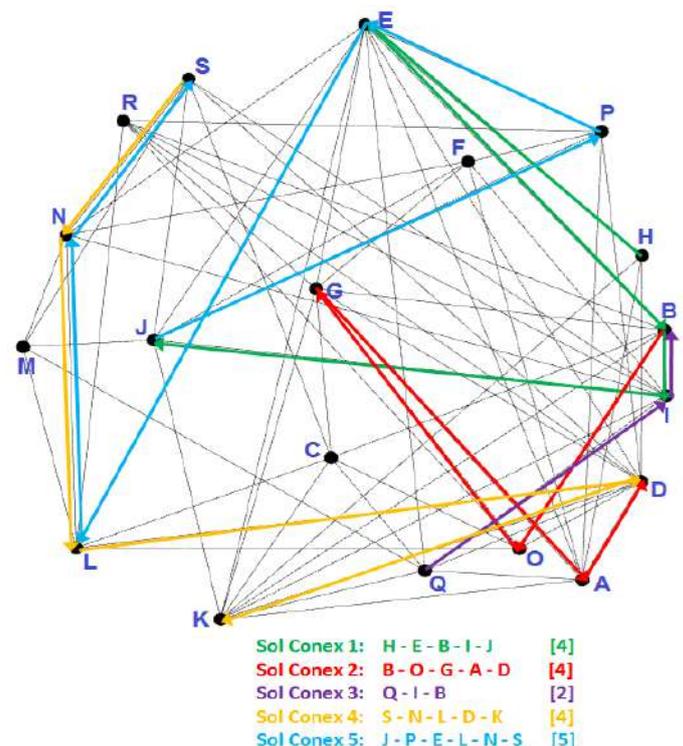


Ilustración 5: Mejor Solución encontrada para la instancia de 19 nodos con 5 conexiones. Coste en número de saltos junto a cada conexión.

N.º de ejecución:	1	2	3	4	5	Solución global 19 nodos
1	4	4	2	4	5	19
2	5	6	2	3	6	22
3	5	6	2	6	7	26
4	4	4	2	3	6	19
5	5	6	2	5	5	23
6	4	6	2	4	6	22
7	5	6	2	5	7	25
8	4	5	2	5	4	20
9	5	5	2	4	5	21
10	4	5	2	5	6	22

Min. coste solución:	19
Peor coste solución:	26
Desv. Típica de todas las soluciones:	2,33
Media de todas las soluciones:	21,9

Tabla 3: datos de las simulaciones para la instancia de 19N y resultados.

N.º de ejecución:	1	2	3	4	5	6	7	8	9	10	Solución global 52 nodos
1	2	3	3	3	4	4	2	4	3	2	30
2	2	4	3	4	3	3	2	4	3	2	30
3	2	3	3	3	4	3	3	3	3	2	29
4	2	5	4	3	3	3	2	4	4	3	33
5	3	5	4	4	3	3	2	3	3	2	32
6	3	4	3	3	4	4	2	3	3	2	31
7	2	4	3	3	3	3	2	4	3	2	29
8	3	4	3	4	3	3	2	4	4	2	32
9	2	3	3	3	4	3	2	3	3	2	28
10	2	4	3	4	4	3	2	4	4	3	33
Min. coste solución:											28
Peor coste solución:											33
Desv. Típica de todas las soluciones:											1,77
Media de todas las soluciones:											30,7

Tabla 4: datos de las simulaciones para la instancia de 52N y resultados.

## VII. CONCLUSIONES

En este trabajo ha sido presentado una adaptación de la metaheurística de Optimización basada en Colonia de Hormigas para la resolución del problema de optimización del routing para SFC (OR-SFC).

Los tiempos medios de ejecución arrojados por las simulaciones de las tres instancias procesadas (6, 19 y 52 nodos) son objeto de interés dado su potencial ofrecido para ejecuciones en tiempo real en redes 5G, si bien como se ha mencionado anteriormente podrían aplicarse restricciones de tiempo sin mayor dificultad al algoritmo.

A partir de estos resultados se procederá a mejorar el algoritmo en algunos aspectos tales como aplicar una heurística más efectiva para guiar la búsqueda hacia nodos que puedan servir la función virtual requerida y aplicar Búsqueda Local para mejorar las soluciones más fácilmente. Además, otra opción sería implementar otro tipo de modelo OCH que pudiese ser más efectivo y poder compararlos, como podrían ser el Sistema de Colonia de Hormigas [9] o el Sistema de Hormigas MAX-MIN [10], obteniendo así una mayor base para poder decantarnos en centrarnos en uno u otro.

## AGRADECIMIENTOS

Este trabajo ha sido realizado como proyecto de Trabajo de Fin de Grado en la Universidad de Granada, con la ayuda de Antonio Miguel Mora García actuando a modo de tutor docente.

## REFERENCIAS

- [1] V. Eramo, E. Miucci, M. Ammar, and F. G. Lavacca, "An approach for service function chain routing and virtual function network instance migration in network function virtualization architectures," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2008–2025, Aug. 2017.
- [2] T. Lukovszki, M. Rost, and S. Schmid, "It's a Match!," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 46, no. 1, pp. 30–36, Jan. 2016.
- [3] D. Sudholt and C. Thyssen, "Running time analysis of ant colony optimization for shortest path problems," *J. Discret. Algorithms*, vol. 10, no. 1, pp. 165–180, Jan. 2012.
- [4] A. M. Mora, J. J. Merelo, P. A. Castillo, and M. G. Arenas, "HCHAC: A family of MOACO algorithms for the resolution of the bi-criteria military unit pathfinding problem," *Comput. Oper. Res.*, vol. 40, no. 6, pp. 1524–1551, Jun. 2013.
- [5] Chandana M. and S. Thakur, "Ant-Net: An adaptive routing algorithm," in *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, 2016, pp. 1–4.
- [6] A. Laghrissi and T. Taleb, "A Survey on the Placement of Virtual Resources and Virtual Network Functions," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 2. Institute of Electrical and Electronics Engineers Inc., pp. 1409–1434, 01-Apr-2019.
- [7] J. Gil Herrera and J. F. Botero, "Resource Allocation in NFV: A Comprehensive Survey," *IEEE Trans. Netw. Serv. Manag.*, vol. 13, no. 3, pp. 518–532, Sep. 2016.
- [8] J. L. Deneubourg, J. M. Pasteels, and J. C. Verhaeghe, "Probabilistic behaviour in ants: A strategy of errors?," *J. Theor. Biol.*, vol. 105, no. 2, pp. 259–271, 1983.
- [9] M. Dorigo and L. M. Gambardella, "Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem," 1997.
- [10] T. Stützle and H. H. Hoos, "MAX-MIN Ant System," *Futur. Gener. Comput. Syst.*, vol. 16, no. 8, pp. 889–914, 2000.



**Segundo Moreno Torres** es Graduado en Ingeniería de Tecnologías de Telecomunicación con mención en Telemática por la Universidad de Granada (junio 2019). Ha colaborado en dos artículos: uno como segundo autor (Composición de Cadenas de Servicios con un Algoritmo de Optimización Basada en Colonias de Hormigas, publicado en JITEL Zaragoza 2019) y otro como primer autor (*Applying Ant Colony Optimization for Service Function Chaining in a 5G Network*, expuesto en el sexto congreso IOTSMS en Granada).



**Antonio M. Mora García** es Doctor en Informática por la Universidad de Granada desde mayo de 2009, donde también obtuvo su Ingeniería Informática en 2001. En la actualidad trabaja como Profesor Contratado Doctor dentro del Departamento de Teoría de la Señal, Telemática y Comunicaciones en la misma universidad. Ha participado en más de 20 proyectos de investigación nacionales e internacionales, tanto como investigador contratado, como colaborador, y ha dirigido 4 proyectos (1 nacional, 1 regional y 2 en la UGR). Tiene más de 20 publicaciones en revistas internacionales indexadas en JCR, y más de 80 trabajos en congresos internacionales de prestigio. Sus áreas de investigación incluyen principalmente la aplicación de algoritmos bioinspirados a diversos problemas reales.

# DESARROLLO DE UN VIDEOJUEGO E IMPLEMENTACIÓN DE TÉCNICAS DE MEJORA DE LA QOE

Autor: Juan Miguel García Martínez, e-mail: juanmgarcia@correo.ugr.es

Tutor: Juan José Ramos Muñoz, e-mail: jjramos@ugr.es

Cotutor: Jonathan Prados Garzón, e-mail jpg@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—Con la llegada de las redes 5G se prevé la definición de nuevas aplicaciones con requisitos de red realmente ajustados. Por ejemplo, las aplicaciones de Internet táctil requerirán un retardo máximo extremo a extremo inferior a 1 ms. Otras aplicaciones, como los juegos en red, pueden tener requisitos de tiempo real similares. Este proyecto se centra en el desarrollo de un videojuego multijugador *online* con requisitos de tiempo real, analizar el impacto del retardo sobre la calidad percibida por los usuarios, y proponer mecanismos que mitiguen este efecto. Para ello se revisa el estado del arte sobre el tema. Después se realiza un análisis sobre la información recogida dando una primera estructura al juego para, posteriormente, realizar un diseño del videojuego y de las técnicas de mitigación aplicables. Se describe después su implementación y, por último, se lleva a cabo una evaluación de los mecanismos propuestos.

**Palabras clave**—Internet táctil, *Mean Opinion Score* (MOS), *Mobile Edge Computing* (MEC), motor de videojuegos, multijugador, *Quality of Experience* (QoE), retardo.

## I. INTRODUCCIÓN

EN este trabajo se desarrolla un videojuego multijugador *online* con necesidades de tiempo real con el motor de videojuegos GameMaker: Studio [1]. Además se implementan mecanismos para mejorar la satisfacción que perciben los jugadores o *Quality of Experience* (QoE) (Calidad de experiencia, en castellano). Una de las razones por las que se desarrolla un juego con requisitos de tiempo real es experimentar con mecanismos que puedan aplicarse también a servicios de Internet táctil. Nótese que las aplicaciones de Internet táctil tienen unos requisitos muy estrictos, necesitando retardos extremo a extremo inferiores a 1ms. Internet táctil supone la siguiente evolución de Internet, y supondrá un gran avance en el mundo de las telecomunicaciones, con la llegada de las redes *Fifth Generation* (5G). Para poder satisfacer los requisitos de retardos de Internet táctil, se están proponiendo técnicas que se aplican en la red, cerca de los usuarios, y que se basan en predecir las acciones de los participantes. Estas ideas son las que se aplican en el juego desarrollado, como se detallará más adelante.

## II. ESTADO DEL ARTE

El estado del arte engloba la situación actual, en el caso de las ingenierías, de las distintas tecnologías existentes y su último desarrollo. Para este trabajo se ha investigado sobre tres áreas.

### A. Internet táctil

Según la *International Telecommunication Union Telecommunication standardization sector* (ITU-T) [2], el concepto de Internet táctil surge como el siguiente paso en la evolución de Internet, donde hoy en día está aflorando el *Internet of Things* (IoT) como una revolución que permite compartir información proveniente de todo tipo de sensores. La Internet táctil requiere de un retardo extremo a extremo de 1 ms, y para hacerla posible [3], es necesario que la tecnología avance en:

- **Redes ultrarrápidas:** Para las exigencias de la Internet táctil son necesarios más cambios en la arquitectura de la red. El 5G jugará un papel esencial para reducir el retardo extremo a extremo.
- **Códecs del tacto:** Códecs que puedan trabajar con el tacto y el movimiento de manera que compriman los datos recogidos antes de transmitirlos por la red.
- **Inteligencia Artificial (IA):** Es necesario predecir la información que transmite el emisor en comunicaciones donde es imposible reducir el retardo. Puede conseguirse sobre todo cuando se tiene periodicidad.

Es necesario destacar la propuesta del *Mobile Edge Computing* (MEC) [4] como aproximación para solventar el problema de los retardos. Consiste en la ejecución parcial de un software en los extremos de la red, con el objetivo de que el usuario obtenga los resultados de algunas funciones más rápido debido que un nodo cercano se los puede proporcionar.

La Internet táctil tendrá un gran impacto en aplicaciones como la telemedicina o la conducción autónoma, y en el ámbito de los videojuegos, donde cada año aparecen juegos que tienen altas necesidades de tiempo real.

### B. Provisión de Calidad de Experiencia (QoE)

La ITU-T [5] define la QoE como el grado de agrado o desagrado para un usuario de una aplicación o servicio. Por ello también se le llama calidad subjetiva. La ITU-T [6] describe que la QoE puede ser medida y aporta diferentes maneras de hacerlo para una aplicación multimedia.

Las medidas de la QoE pueden realizarse de manera subjetiva y objetiva, las cuales dan resultados en una escala denominada *Mean Opinion Score* (MOS), siendo 1 una mala

experiencia, 2 mediocre, 3 aceptable, 4 buena y 5 una calidad excelente.

Las medidas subjetivas pueden realizarse mediante los siguientes procedimientos, descritos en [6]: *Índices por categorías absolutas*, *Índice por categoría absoluta con referencia escondida*, *Índices por categorías de degradación*, *Método de comparación por pares*, *Doble estímulo simultáneo para una evaluación continua* y *Evaluación por objetos*.

En este proyecto se han utilizado dos de ellos:

- **Índices por categorías absolutas:** Se muestra al usuario una secuencia, y cuando finalice, el usuario la evalúa con la escala MOS. Después se pasa a la siguiente secuencia.
- **Evaluación por objetos:** Consiste en que el usuario evalúe primero una escena con un fondo y objetos, y después evaluar solamente los objetos por separado con la escala MOS.

Por otro lado, las medidas objetivas se consiguen mediante modelos matemáticos que relacionen el MOS con otros factores, los más representativos son el *delay* y el *jitter*.

Para mejorar la QoE existen distintas técnicas que se adecuan mejor o pero dependiendo de la aplicación que se pretende mejorar. En el apartado de análisis se muestran las técnicas implementadas en este proyecto.

### C. Motores de videojuegos

Existe una gran variedad de motores de videojuegos con diferentes funcionalidades que deben ser estudiadas con el fin de comprobar si son aptos para el juego que se quiere crear.

Estas herramientas comprenden un conjunto de utilidades que permiten el desarrollo de videojuegos a gente sin requerir unos altos niveles de programación, como era el caso de los juegos antiguos, por lo tanto, que una herramienta sea fácil de utilizar da lugar a que más gente pueda utilizarla. Algunas de ellos son *Unity*, *Unreal Engine 4*, *CryEngine V*, *GameMaker Studio*, *Amazon Lumberyard* o *JMonkey*.

En el trabajo se ha utilizado *GameMaker Studio* [1] que utiliza un lenguaje propio basado en C pero orientado a objetos. Permite crear videojuegos multiplataforma y también videojuegos con conectividad *online* utilizando *sockets* Transmission Control Protocol (TCP) y User Datagram Protocol (UDP), por lo que deja al usuario una configuración amplia a la hora de diseñar los mensajes a enviar y qué respuesta da el receptor a estos mensajes.

## III. ANÁLISIS

A continuación se procede a un análisis de la información recopilada en el estado del arte para extraer los datos más importantes y tomar decisiones sobre el desarrollo del proyecto.

### A. Aproximando la Internet táctil

La red con esas características ultra rápidas aún no está implantada y para el juego no serán necesarios unos códecs del tacto, por otro lado, la IA juega un importante papel a la hora de reducir la latencia ya que permite realizar predicciones en caso de que el paquete con la información se retrase. Asimismo, el concepto de MEC ya utilizado en otros juegos permite evitar transmitir información por la red que puede ser obtenida de manera local, con lo que el efecto del *delay* es mucho menor.

### B. Técnicas de mejora de QoE que se van a implementar

Para separar las distintas partes que pueden afectar a la QoE se ha realizado una división entre métodos donde la importancia radica en la red y métodos donde se prioriza el diseño o características del videojuego en sí. *Mejoras ante los problemas de red:* En el caso de las mejoras donde tiene mayor relevancia la propia red, se presentan en la siguiente lista aquellas mejoras escogidas:

- **Predicción:** La predicción es uno de los factores necesarios para que se consiga Internet táctil, por ello se implementará en este proyecto que pretende seguir los principios de la Internet táctil, realizándose dos tipos de predicción, una que ayude al jugador más retardado y otra para mitigar la pérdida de paquetes o el alto retardo.
- **Uso del MEC:** De la misma manera, el MEC es una tecnología necesaria para conseguir Internet táctil, por lo tanto debe haber un evento del juego que aunque a primera impresión lo necesario sea enviar un mensaje al servidor sobre el evento ocurrido, puede evitarse el envío de dicho mensaje y operar localmente, tanto si el evento le ocurre al jugador local como si le ocurre a uno *online*.
- **Elección correcta del servidor:** En caso de disponer de más de un servidor, es importante que el jugador sepa a cuál de ellos debe conectarse, por lo que se implementa algún método para hacerle llegar esta información al jugador.

*Mejoras respecto al diseño del juego:* Para el caso de mejoras que recaen sobre el diseño del juego, se muestran a continuación las que se introducirán en el juego:

- **Matchmaking:** Se incluye algún método que implique la separación de jugadores novatos y expertos. Se implementará porque de manera rápida y con pocas partidas, los jugadores tendrán una mejor experiencia al jugar con otros de habilidades similares.
- **Distintos personajes:** Se ofrecen distintos personajes para que el jugador escoja aquel que más le satisfaga. Es una manera sencilla que requiere mayormente de un diseño atractivo para que el jugador juegue más cómodamente.
- **Comunicación:** En los videojuegos online es muy común el poder enviar mensajes de un jugador al resto ya que ofrece una experiencia más completa, por este motivo, se añaden elementos que permitan a los jugadores comunicarse entre sí.

### C. Método de evaluación

Para evaluar la QoE, esto es, obtener el MOS de los usuarios tras probar el juego, se recurren a los tests de opinión puesto que se quiere una medida subjetiva en base a la experiencia, por lo que en primer lugar se eligen a los *beta testers* que examinarán el juego y se comprueba si las técnicas implementadas mejoran realmente su QoE.

Se utiliza el método de evaluación por objetos para evaluar tanto aspectos concretos como el contenido general, en los cuales también se hace uso del método de índices por categorías absolutas.

Adicionalmente, se calcula una función que relaciona retardo y MOS para poder estimar la QoE que reciben los jugadores con un retardo dado.

#### D. Elección del motor del videojuego

Se ha escogido GameMaker Studio ya que dispone de una interfaz visual y sencilla con la que se puede trabajar cómodamente. Permite la integración de online en el juego con ajustes avanzados, y dispone de una gran documentación con respecto a los otros motores estudiados. La documentación por parte de la propia empresa y por los foros es muy amplia y útil.

### IV. DISEÑO DE LA SOLUCIÓN

En el diseño es donde se ha escogido aspectos del juego sobre su jugabilidad, los elementos que lo componen y como se implementan los mecanismos de mejora de la QoE así como los protocolos que permiten la conectividad entre jugadores y servidores.

#### A. Descripción del juego

El juego consiste en un juego de plataformas en el que hay que controlar a un personaje, que podrá moverse a la izquierda o a la derecha y hacia arriba o hacia abajo mediante saltos o dejándose caer. Tras estos movimientos básicos, el personaje también podrá teletransportarse una corta distancia cada dos segundos con el objetivo de esquivar proyectiles pero sin abusar de la habilidad. También puede disparar proyectiles en cualquier dirección, apuntando con el ratón. El objetivo del juego es eliminar al resto de personajes quitándole su vida. Puede verse una partida en la figura 1.

#### B. Elementos del juego

Cuando el usuario inicie el juego, se muestra por pantalla un menú donde el jugador puede conectarse a un servidor concreto o realizar una búsqueda de servidores pudiendo escoger el de menor retardo. Una vez conectado a un servidor, puede registrarse o iniciar sesión mediante un nombre y una contraseña, con ello se mantendrá un registro de los jugadores. A continuación, otro menú donde comenzar la búsqueda de una partida y escoger su personaje.

Para implementar todos los elementos se han creado clases que interactúan entre sí para permitir la ejecución del juego, las clases se muestran en la figura 2. La clase principal es *obj\_controlador*, ya que es la que crea y mantiene la conexión con el servidor, permitiendo también que se envíen todos los mensajes de los protocolos y manejar la comunicación de las demás clases.

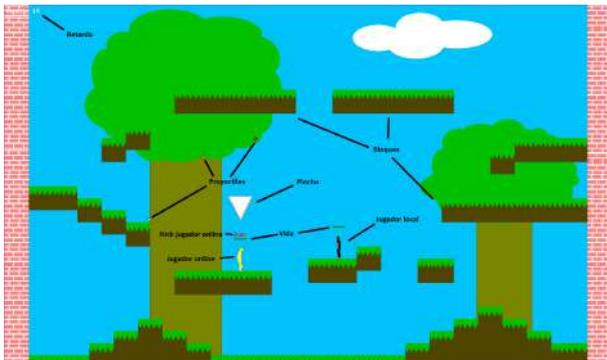


Fig. 1. Vista de una partida.

#### C. Técnicas de mejora de la QoE

Es importante mencionar que conocer el retardo de los jugadores es importante para las mejoras, por lo que se enviarán mensajes encargados de calcular este retardo de manera continuada utilizando una marca de tiempo que se mandará al servidor, después este se la devolverá al jugador y podrá comparar esa marca con el tiempo actual, obteniendo así el retardo.

Las técnicas implementadas son las siguientes:

- **Protocolos de transporte:** Aunque la elección del protocolo de transporte pueda no considerarse un mecanismo de mejora, las diferencias que suponen escoger entre uno u otro pueden comprometer la QoE, por lo que debe tomarse en consideración su elección. En el apartado IV-D se profundiza en este aspecto.
- **Predicción:** Se ha escogido realizar dos tipos de predicciones: Predicción de movimiento y predicción de comportamiento.

La predicción de movimiento afecta al movimiento de los personajes y los proyectiles. En caso de pérdida de paquetes o alto retardo, se traza una recta imaginaria tomando las posiciones anteriores que ocupaban estos elementos para predecir la posición que ocuparían en los instantes de tiempo posteriores.

Cuando un jugador experimenta un gran retardo, los movimientos de este puede que lleguen demasiado tarde al resto de jugadores. De esta forma, aunque el jugador fuera a esquivar un proyectil, no tendría la oportunidad de hacerlo debido a su retardo, y se contaría un impacto que pudo no haber ocurrido.

Normalmente, cuando un proyectil del jugador local impacta en un enemigo, el jugador local se lo notifica al servidor. Sin embargo, para mitigar el problema descrito se hace lo siguiente: Si el jugador retardado tiene un *delay* de 500 ms, retardo muy alto para el juego, es este quien notifica al servidor si es impactado por un proyectil. Además, el resto de usuarios ven que el jugador retardado esquiva los proyectiles para evitar que observen cómo es impactado sin perder vida.

- **Ejecución local de un evento online:** Cuando un jugador esta quieto unos segundos, cae un pincho encima suya, y sería lógico que este evento fuese generado por el servidor. Sin embargo, esto requeriría el envío de men-

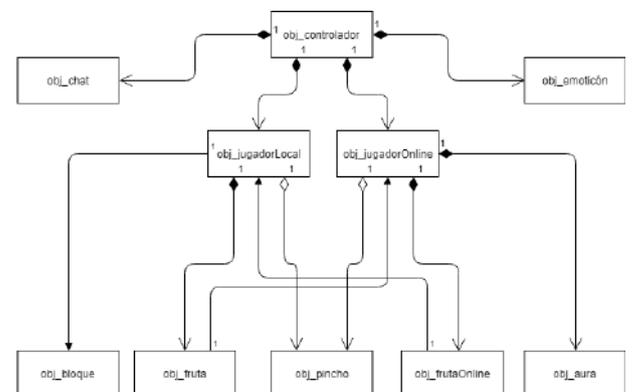


Fig. 2. Diagrama *Unified Modeling Language* (UML) de las clases del personaje, proyectiles y elementos de comunicación.

sajes para ejecutar acciones que podrían ser ejecutadas localmente sin necesidad de un paquete proveniente del servidor. Para solucionarlo, el cliente hace aparecer el pincho por su cuenta siguiendo la filosofía de la Internet táctil sobre el MEC. Para que los otros jugadores creen el pincho al mismo tiempo, se ajusta según su retardo.

- **Búsqueda de servidores:** La elección del servidor adecuado es un factor que afecta positivamente a la QoE ya que repercute en el retardo del usuario. Para que un nuevo jugador conozca a que servidor debe conectarse, se dispone de otro servidor distinto, denominado Servidor maestro, que almacena los servidores que existen, así, los jugadores pueden solicitar la lista de servidores y comprobar el retardo con cada uno de ellos para conocer al servidor al que deberían conectarse en el futuro.
- **Emparejamiento (Matchmaking):** A la hora de solicitar partida, el servidor junta a los jugadores de mayor nivel en una partida y a los de menor en otra. Este registro se mantiene aunque se apague el juego o el servidor para que el jugador no empiece desde el nivel mínimo cada vez que arranca el juego.
- **Elección de personajes:** Siguiendo un patrón de movimientos y posturas, se tienen 4 diseños distintos de personajes. Así, los jugadores pueden escoger un personaje antes de comenzar la partida y ofrecer una experiencia más personalizada.
- **Comunicación entre jugadores:** Para que los jugadores se puedan comunicar entre sí, se dispone de un chat para enviar textos. Además, se pueden enviar de forma ágil emoticonos que expresen felicidad o enfado durante la partida.

un servidor el cual se identifica como tal, el Servidor maestro le otorga un número de identificación, lo añade a una lista de servidores disponibles y mantiene su conexión activa. Cuando el servidor se desconecta, el maestro lo borra de la lista de servidores disponibles.

- **Conexión Jugador a Servidor maestro:** Un jugador se conecta al Servidor maestro, solicitando la lista de servidores disponibles, cuando los ha recibido todos el Servidor maestro el indica que se puede desconectar y el jugador lo hace.
- **Prueba de retardo:** Una vez que el jugador posee la lista de servidores disponibles, se realiza una prueba para comprobar cuales de los servidores ofrece un menor retardo, este planteamiento simula cuando un juego tiene servidores en distintas localizaciones y según la residen-

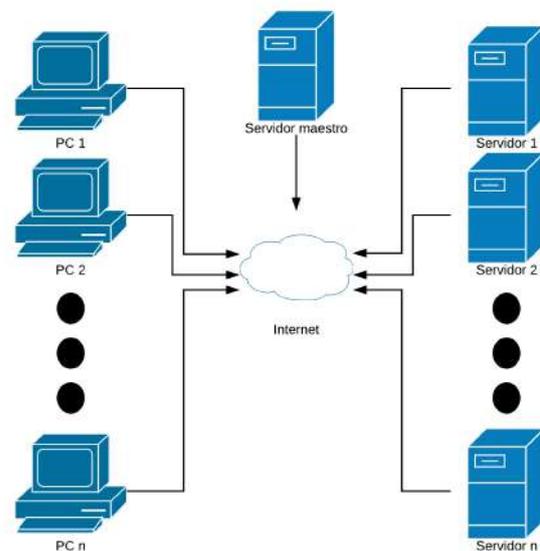


Fig. 3. Modelo de conexión cliente-servidor del juego.

#### D. Diseño del protocolo de red

Para llevar a cabo la conexión entre los jugadores y el servidor se hace uso del protocolo de red *Internet Protocol version 4* (IPv4) y de los protocolos de transporte tanto *Transmission Control Protocol* (TCP) como *User Datagram Protocol* (UDP).

Para el establecimiento de conexiones, inicio de sesión, registro y algunos aspectos del desarrollo del juego se utiliza TCP ya que es un protocolo con medidas para contrarrestar la pérdida de paquetes, por lo que su uso se limita a aquellos mensajes que son necesarios enviar sin errores ni pérdidas. Por otra parte, para el movimiento de los personajes, sus proyectiles y el cálculo del retardo se hace uso de UDP ya que se tratan de paquetes en los que es más importante que el retardo sea bajo y la pérdida de paquetes no es tan importante, como suele ocurrir en aplicaciones multimedia y UDP ofrece un menor retardo que TCP. Por otro lado, como protocolo de aplicación se ha diseñado uno específico para el videojuego permitiendo la conexión de los jugadores entre sí mediante un servidor.

En la figura 3 puede verse la topología de red del juego, con lo que se pueden distinguir 4 tipos de conexiones, y en la figura 4 el diagrama de estados de la parte del protocolo encargada de comunicar al Servidor maestro con el jugador y servidor.

- **Conexión Servidor a Servidor maestro:** El Servidor maestro está escuchando hasta que recibe la conexión de

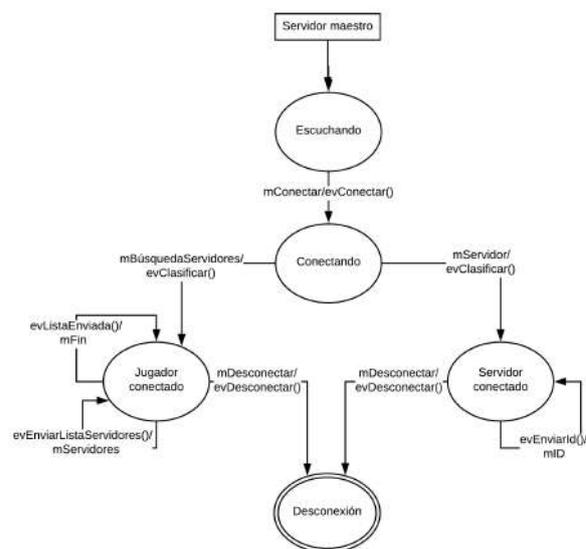


Fig. 4. Diagrama de estados del Servidor maestro.

cia del jugador tendrá menor retardo unos que otros.

- **Conexión normal:** Un jugador pretende jugar, al conectarse al servidor, este le proporciona un número de identificación y procede a registrarse o iniciar sesión, una vez hecho, el jugador ya puede comenzar una partida *Player versus Player* (PvP) donde se esperan a los demás jugadores y se les asigna a cada uno una partida distinta según su nivel con el fin de juntar a rivales de habilidades similares. Una vez se reúnen los jugadores, comienza la partida, y al finalizar, se actualizan los niveles de los jugadores.

### E. Mensajes

Los mensajes diseñados para el protocolo de comunicación del juego tienen una longitud variable, múltiplo de 8 bits, dependiendo del tipo que sea.

Todos los mensajes comienzan con un capo denominado *tipoPaquete* que determina cómo el receptor debe interpretar dicho mensaje. En el caso de una conexión normal, se añade otro campo al final de cada mensaje denominado *númeroPaquete*, el cual se utiliza como número de secuencia para que el servidor sepa el orden de los mensajes. Esto se incluye para que el protocolo esté preparado para una mejora futura del juego donde el servidor pueda detectar posibles trampas en los jugadores enviando paquetes en distinto orden o para calcular la tasa de pérdida de paquetes. Este campo se incluye también en el transcurso de la partida.

## V. IMPLEMENTACIÓN

En esta sección se explica cómo se ha llevado a cabo el desarrollo del videojuego siguiendo las pautas del diseño creando los *scripts* y las clases necesarias.

Para la transmisión de mensajes, los servidores y los clientes utilizan principalmente una clase *obj\_controlador* y un *script scr\_recibiendoPaquetes*. En la clase mencionada se crea un *buffer* donde guardar la información que se quiere enviar y la que se recibe. Además se crean los *sockets* TCP y UDP que se van a utilizar. A continuación, la máquina que actúe de servidor lo crea, y se procede a la conexión entre cliente y servidor.

### A. Mejoras

En este apartado se explica la forma en la que se han implementado las mejoras en el videojuego, por lo que se detalla a continuación siguiendo el orden del apartado IV, diseño:

- **Protocolos de transporte:** Se han creado dos *sockets* en los servidores, uno TCP y el otro UDP. De esta manera, los clientes envían mediante *socket* TCP todos los mensajes salvo el movimiento de los jugadores, proyectiles, y paquetes de cálculo del retardo que se envían con UDP.
- **Predicción:** Para la predicción de movimiento, en primer lugar, se guardan las posiciones que va ocupando el proyectil, y si hay al menos 3, se comprueba si las posiciones actuales son iguales a las anteriores, lo que indicaría que está quieta por lo que se han podido perder paquetes. Por lo tanto, se traza una recta con esos dos

puntos y se le asigna la nueva posición que debe ocupar. Para el movimiento de los personajes es similar pero se comprueba que está si estaba en movimiento ya que si el jugador ha hecho que su personaje se quede quieto no es necesario realizar ninguna predicción.

Por otro lado, para la predicción de comportamiento, se ha creado un aura que rodea a los personajes enemigos, y si su retardo está entre 0.5 y 1 segundo, al chocar con un proyectil lo hace esquivar. Luego se verifica si efectivamente el jugador había esquivado o no el proyectil, ajustando su vida.

- **Ejecución local de un evento *online*:** Para evitar la necesidad de enviar un mensaje a la hora de crear un pincho, el cliente tiene que hacerlo automáticamente para seguir la filosofía MEC, para ello, se comprueba si el personaje está quieto unos segundos, ese tiempo se ha tomado como 2 segundos menos el retardo que tenga el personaje con el fin de equiparar el momento de caída de un pincho tanto para el jugador local como para los jugadores *online*.
- **Búsqueda de servidores:** Para hacer una búsqueda de servidores, es preciso que existan servidores conectados al servidor maestro para que una vez que el jugador le consulte, pueda obtener una lista con los servidores activos. En cuanto se realiza la conexión jugador-servidor maestro, este último le envía la lista en mensajes como los siguientes, donde un mensaje indica también que se ha enviado la lista completa. Una vez recibida, el jugador comienza la prueba de retardo para comprobar a que servidor debe conectarse.
- **Emparejamiento (*Matchmaking*):** Para realizar la clasificación de jugadores, estos al acabar una partida, actualizan su nivel, aumento en mayor medida si son los últimos supervivientes. En el servidor se crea una matriz que contiene a los jugadores esperando una partida con su nivel, después se ordena de mayor nivel a menor y se les asigna la partida o *room* difícil a los 3 primeros y la fácil a los 3 siguientes, ya que las partidas son de 3 jugadores. Cada vez que alguien quiere empezar una partida, el servidor les envía un mensaje como el siguiente a los jugadores informando sobre la *room* a la que van a ir.
- **Elección de personajes:** Antes de comenzar una partida, el jugador tiene la posibilidad de escoger entre 4 personajes clickando en el que prefieran utilizar.
- **Comunicación:** La comunicación se ha desarrollado mediante un intercambio de mensajes donde un jugador envía un mensaje al servidor y el resto de jugadores lo reciben mostrándose encima del jugador emisor. De manera similar, se pueden enviar también emoticonos.

### B. Tipos de conexiones

Para la implementación de los clientes y servidores se ha tenido que seguir el siguiente orden en referencia a los tipos de conexiones explicados en el diseño:

- **Conexión Servidor a Servidor maestro:** Se crea al Servidor maestro abriendo un *socket* TCP en el puerto 49999. A continuación, un servidor realiza la conexión

con el Servidor maestro indicando la dirección IP, el puerto, y de que se trata de una conexión TCP.

Con esto el Servidor maestro registra al servidor conectado y le da un número de identificación.

- **Conexión Jugador a Servidor maestro:** La conexión es similar a la anterior, pero en el campo *tipoPaquete* del mensaje se indica que se trata de un jugador.
- **Prueba de retardo:** Los servidores abren un puerto UDP en el 50001, y después, el jugador envía los paquetes para medir el tiempo a los servidores mediante UDP a ese puerto y calcula el retardo con todos ellos.
- **Conexión normal:** Para la conexión normal y hacer posible las partidas, el servidor debe crear un *socket* TCP en el puerto 50000 donde los jugadores realizan la conexión, se registran y comienzan sus partidas. Sin embargo, los movimientos de los jugadores y proyectiles se envían mediante mensajes UDP como parte de las mejoras respecto al retardo.

## VI. PRUEBAS

Para realizar las pruebas se ha dispuesto de distintas topologías basadas en el modelo ideal de la figura 3 dependiendo del aspecto concreto a experimentar, sin embargo, tienen en común que se utiliza un *router* y varios ordenadores que hará el papel de servidor, servidor maestro o cliente.

Las pruebas determinaron que las técnicas introducidas fueron de ayuda al mejorar la QoE, sin embargo, los dos más relevantes se desarrollan a continuación.

La figura 5 muestra como al haber poco retardo, una versión del juego sin mejorar y otra mejorada no presentan apenas diferencia respecto a la QoE, sin embargo, a medida que aumenta el retardo, se puede apreciar que la versión mejorada está dando mejores resultados que la versión básica, al menos, hasta un retardo de 500 ms aproximadamente donde el retardo ya es demasiado grande como para que la diferencia sea notable.

De esta gráfica puede extraerse la ecuación que da una medida objetiva de la QoE respecto al retardo en este videojuego.

$$QoE_{enhanced}(x) = 0.00000003X^3 - 0.00002X^2 - 0.0026X + 4.4788 \quad (1)$$

Por otro lado, también se ha podido medir la relación entre QoE y el porcentaje de pérdida de paquetes, donde se obtuvo que con un porcentaje de pérdida de paquetes pequeño, de un 2%, la versión mejorada ayuda a mantener la QoE en una posición alta. A medida que se incrementa la pérdida

de paquetes, ambas versiones, mejorada y básica, acaban convergiendo, equivaliendo la QoE con un 30% de pérdida de paquetes.

## VII. CONCLUSIONES

La creación de un videojuego supone un trabajo duro pero gratificante, y hoy día gracias a la multitud de motores que existen en el mercado y la cantidad de información en la red, resulta muy accesible a todo el mundo que quiera probarlo. Y por eso mismo, para conseguir un producto que destaque sobre los demás, tiene que resultar muy gratificante para el usuario, por lo que la inclusión de técnicas para mejorar la QoE es vital.

En este proyecto, se han abordado diferentes formas de incrementar la QoE, y se ha probado mediante una fase de experimentación que resultan realmente útiles, por lo que este campo es interesante para continuar estudiándolo durante los próximos años.

## AGRADECIMIENTOS

A mis padres por preocuparse por mí a lo largo de toda la carrera, a mi hermana por abrirme tantas puertas, a mi hermano por ayudarme con cualquier problema y por insistirme en lo útil que es un *debugger*, a los compañeros que me han acompañado a lo largo de la carrera por lo que hemos vivido dentro y fuera de ella, a mis amigos que se ofrecieron a realizar las pruebas que requieren este proyecto y a Juanjo y Jonathan por aconsejarme y encontrar siempre un hueco para las tutorías.

## REFERENCIAS

- [1] "Yoyo games," Jun. 2018. [Online]. Disponible: <https://www.yoyogames.com/>
- [2] ITU-T, "The tactile internet," *ITU-T Technology Watch Report*, Aug. 2014. [Online]. Disponible: <http://www.itu.int/en/ITU-T/techwatch/Pages/tactile-internet.aspx>
- [3] M. Dohler, "Global reach: Will the tactile internet globalize your skill set?" *IEEE ComSoc Technology News*, Jan. 2017. [Online]. Disponible: <https://www.comsoc.org/ctn/global-reach-will-tactile-internet-globalize-your-skill-set>
- [4] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing a key technology towards 5g," *ETSI White Paper No. 11*, Sep. 2015. [Online]. Disponible: [http://www.etsi.org/images/files/ETSIWhitePapers/etsi\\_wp11\\_mec\\_a\\_key\\_technology\\_towards\\_5g.pdf](http://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp11_mec_a_key_technology_towards_5g.pdf)
- [5] ITU-T, "New definitions for inclusion in recommendation itut p.10/g.100recommendation itu-t p.10/g.100(2006)-amendment5," *Recommendation P.10/G.100 Amendment 5*, Sep. 2016. [Online]. Disponible: <https://www.itu.int/rec/T-REC-P.10-201607-S!Amd5/en>
- [6] ITU-T, "Subjective video quality assessment methods for multimedia applications," *ITU-T Study Group 9*, Apr. 2008. [Online]. Disponible: <https://www.itu.int/rec/T-REC-P.910-200804-I/en>

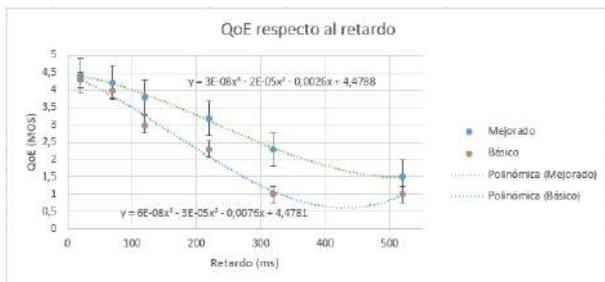


Fig. 5. QoE respecto al retardo.



**Juan Miguel García Martínez**, nacido el 21 de Octubre de 1995 en Granada y graduado en Ingeniería de Tecnologías de Telecomunicación por la Universidad de Granada.

# Diseño e implementación de un videojuego didáctico multijugador

Autor: Pedro José Arias Perandrés; e-mail: pedroarias@correo.ugr.es

Tutor: Juan José Ramos Muñoz; e-mail: jjramos@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones

Universidad de Granada

**Resumen**— El presente trabajo de fin de grado tratará sobre el diseño y desarrollo de un videojuego multijugador de carácter didáctico con el uso de un motor de desarrollo profesional. Para contextualizar todo el proyecto, se hablará sobre los motores de desarrollo en la actualidad, haciendo un análisis en profundidad del motor de desarrollo utilizado. Asimismo, se ofrecerá una visión general y un análisis en profundidad del framework de red del motor y sus protocolos asociados.

En el trabajo se presenta todo el proceso de desarrollo del videojuego didáctico, desde el diseño de los contenidos, creación de los personajes y escenarios, hasta la programación de la lógica del juego y el soporte multijugador en red. Para ello, se detallará la forma de trabajo en el entorno seleccionado.

Por último, se realizará una caracterización del tráfico de red generado por el presente juego, además de analizar el tráfico de red generado por cada uno de los eventos del mismo y se identificarán optimizaciones y mejoras tanto implementadas como futuras.

**Palabras clave**— Unreal Engine 4, desarrollo de videojuegos, docencia, aprendizaje, red, multijugador, análisis de tráfico.

## I. MOTIVACIÓN

El sector de los videojuegos es uno de los que mayor facturación generan dentro de la industria digital, con un enorme crecimiento en los últimos años como pueden demostrar numerosos estudios, tales como [1], [2] y [3]. El uso de las tecnologías como Internet ha propiciado un cambio evolutivo en la forma que se tiene de ver y jugar a videojuegos en los últimos años, de forma que hemos pasado de jugar a un juego físico en una videoconsola dentro de casa a poder jugar desde cualquier parte del mundo y compartir estas experiencias, puntuaciones y logros con personas de todo el mundo.

Tan solo en España este sector equivale al 0,11% del Producto Interior Bruto (PIB) del país y emplea directamente a 8.790 personas, generando un impacto de 3.577 millones de euros según un informe económico elaborado por la Asociación Española de Videojuegos, véase [4]. A nivel mundial las cifras son aún más espectaculares llegando a alcanzar los 108.9 billones de dólares en 2017 como muestra la Ilustración 1. Hablamos por tanto de que no sólo programadores, sino también artistas, compositores, guionistas, personas de marketing e incluso actores son algunos de los puestos de trabajo involucrados en la creación de un videojuego, véase [5].

La aplicación de videojuegos a nivel docente se ha convertido en una metodología muy empleada a lo largo de los años de forma que se pueda formar a un alumno, pero de forma entretenida y sin que ésta pueda ser vista como una obligación, tal y como demuestran estudios como [6].

## II. OBJETIVOS DEL PROYECTO

Los dos objetivos principales sobre los que se construye el presente trabajo son:

A. *Desarrollar un videojuego multijugador didáctico.*

Desarrollar un videojuego con enfoque multijugador y que además cuente con contenidos didácticos, de forma que se pueda entender como una tarea dirigida a alumnos para que complementen la formación educativa en historia y cultura, acercándolos de una forma más dinámica y divertida al aprendizaje.

B. *Analizar y caracterizar el tráfico de red de un motor de juegos profesional.*

Se realizará un videojuego multijugador, lo que implica cierto consumo de red, el presente proyecto estudiará y analizará el estado de la gestión e implementación a nivel de red de un videojuego de esta índole con un motor de desarrollo profesional.

Con estos objetivos en mente, podemos resumir el objetivo principal en desarrollar un videojuego de carácter cooperativo multijugador, en el que los usuarios tengan que resolver una serie retos en forma de puzzles. Los cuales ayudarán a avanzar por los niveles y aprender conceptos de historia y cultura.

## III. ESTADO DEL ARTE

En la actualidad existen un gran número de títulos cuya finalidad no sólo es entretener sino acercar la cultura e historia a los alumnos, la cual se puede hacer de dos maneras, indirecta en la que el aprendizaje se ve como un complemento a la historia principal, o de manera directa de forma que el aprendizaje de conceptos es el hecho fundamental del juego en sí. Podemos ver cada modalidad mediante ejemplos:

A. *Assassin's Creed [7].*

Esta saga se caracteriza por ser una de las más conocidas en el mundo de los videojuegos, además de por su elevada cantidad de ingresos llegando incluso a alcanzar los 96 millones de dólares según fuentes de la propia Ubisoft y uno de los mejores ejemplos de cómo se puede aunar el aprendizaje de historia, cultura y religión en un videojuego de forma entretenida y adictiva.

B. *Libertus: Ruta hacia la Libertad [8]*

Es una aventura gráfica de misterio, ingenio y acción a través de enclaves históricos de 10 ciudades andaluzas en la antigüedad, desarrollado por Karmo Multimedia, una empresa Sevillana que ofrece este título de forma gratuita para su descarga.

#### IV. PLANIFICACIÓN Y ESTIMACIÓN DE COSTES

El proyecto se planteó como TFG para el curso 2017-2018. No obstante, la idea del mismo se planteó en el anterior curso. El proceso de aprendizaje de un motor de desarrollo como es el que se manejará en la realización del mismo ha sido de 6 meses y la realización del videojuego en sí además de documentación y finalización del mismo de 8 meses. El proceso de aprendizaje ha sido el más largo del desarrollo puesto que se ha empezado desde cero. Además, han sido necesarios una serie de sprints, entendidos como temporadas en las cuales se ha dedicado mayor tiempo a la realización del proyecto y se han resuelto mayor número de tareas. La metodología que se ha seguido a la hora de plantear y realizar el trabajo ha sido:

- A. *Aprendizaje de conceptos y términos generales en Unreal Engine 4.*
- B. *Planteamiento de las mecánicas y estilo del videojuego.*
- C. *Aprendizaje de los conceptos y técnicas de programación a nivel de red.*
- D. *Desarrollo del videojuego y sus mecánicas.*
- E. *Adaptación de las mecánicas para el entorno multijugador en red.*

Se puede ver un desglose de las tareas de los sprints finales en la Tabla I. Además de una planificación temporal se han hecho estimaciones de costes de hardware (Tabla II), software (Tabla III).

Mes	Fechas	Duración (Semanas)	Objetivos
Marzo	05/03/18 18/03/18	2	Terminar el primer mapa jugable (nivel 1), diseño y planteamiento de los puzles y las notas que van a empezar a descubrir la historia.
Marzo/Abril	19/03/18 01/04/18	2	Programar la jugabilidad y el objetivo del juego, concretamente el principio del juego, desarrollo de objetivos y resolución del primer mapa.
Abril	02/04/18 08/04/18	1	Revisión del juego, pulido de los menús y comprobar que el desarrollo del juego está bien planteado.
Abril	09/04/18 22/04/18	2	Análisis del tráfico, planteamiento de distintos tipos de casuísticas, revisión del sistema, mejoras y experiencias de usuario.
Abril/Mayo	23/04/18 06/05/18	2	Terminar la documentación. Añadir todas las funciones y estructura planteada.
Mayo	07/05/18 18/05/18	2	Semanas extra para pulir aspectos, añadir elementos y terminar documentación.
Mayo	18 de Mayo	X	Fecha límite para entrega de memoria al tutor.

Tabla I. *Sprint final.*

Tipo de Componente	Componente	Coste
Procesador	Intel Core i7-8700K 3.7Ghz BOX	356 €
Placa base	MSI Z370 Gaming Plus	130 €
RAM	Corsair Vengeance LPX DDR4 3200 16GB	200 €
Ventilador	Noctua NH-U12S	60 €
Fuente de alimentación	EVGA GQ 650W 80 Plus bronze	79 €
Disco duro 1	SSD Samsung 850 EVO 250GB	90 €
Disco duro 2	HDD 1TB Seagate	46 €
Tarjeta grafica	MSI GTX 1060 6GB	325 €
Caja	Phantecks P400	75 €
<b>Total.</b>		<b>1.361 €</b>

Tabla II. *Coste hardware.*

Programa	Licencia	Coste
Unreal Engine 4	Estándar	Gratuito
3DS Max Studio	1 Año	1984.40€ <sup>1</sup>
Microsoft Office	Básica	100€
Blender	Estándar	Gratuito
Adobe Photoshop	Básica	290€
Adobe Mixamo	Estándar	Gratuito
Adobe Fuse	Estándar	Gratuito
Windows 10	Home Edition	145€
<b>Total</b>		<b>535€</b>

Tabla III. *Coste Software.*

El coste de los servicios prestados se estima en 1200€ al mes de media. Dentro de estos servicios podemos generar la siguiente división de 1103 horas totales:

- Programación/gestión de red – 60% lo que equivale a unas 662 horas.
- Artista/Diseñador – 30% - lo que equivale a 331 horas.
- Documentación y gestiones – 10% - lo que equivale a 110 horas.

Coste total de servicios: 9600€

Coste total del proyecto: 14780€

#### V. MOTORES DE DESARROLLO

Existen diversas formas de desarrollar un videojuego. Se puede hacer desde cero programando todos los aspectos del mismo, utilizar bibliotecas que proporcionen las funciones típicas necesarias para un juego, o usar un motor de desarrollo que tenga las herramientas creadas y listas para ser usadas. La primera técnica no es la más común dada la complejidad y el esfuerzo de las mismas, de forma que lo usual es el uso de la tercera alternativa. Teniendo esta filosofía clara, se va a utilizar un motor de desarrollo para crear el videojuego.

Los motores de desarrollo sustentan las bases para producir un videojuego. Un ejemplo de ello puede ser la definición de las reglas de la física (los objetos sufren interacciones entre sí), gravedad (los objetos caen), iluminación (existencia de fotones simulados que generan luces) y sombras (interacción de un objeto ante los fotones). En la actualidad existen una gran cantidad de motores de desarrollo, desde los más profesionales como son el caso de Unity 3d [9], Cryengine V [10] o Unreal Engine 4 [11] a los menos profesionales como pueden ser Godot o RPG maker. En este caso se ha optado por Unreal Engine 4 dada su potencia gráfica y enfoque hacia el modelado 3D, factores importantes a la hora de generar niveles de carácter arquitectónico y dada la posibilidad de usar subsistemas sobre plataformas como Facebook o Steam.

Aunque su curva de aprendizaje es la mayor, lo cual supone un reto extra a la hora del desarrollo del proyecto, pero al ser uno de los motores 3D más profesionales, el esfuerzo es recompensado con un acabado de aspecto de calidad avanzada, como si de un estudio de videojuegos se tratase.

#### VI. UNREAL ENGINE 4

El motor de Unreal Engine 4 es uno de los más profesionales en la actualidad, lo que repercute especialmente en sus funcionalidades. Dichas funcionalidades no se limitan sólo a la creación de las escenas donde se van a mover los personajes sino también cuenta con herramientas de diseño para materiales los cuales darán color, textura y profundidad a todos los objetos que existan, contamos también con herramientas de modelado de huesos para personajes e incluso generadores de colisiones entre otras muchas utilidades.

En términos de programación, existen dos lenguajes con los que se puede programar, el scripting visual también denominado sistema de Blueprints y C++.

En el caso del desarrollo del videojuego se ha tomado la decisión de optar por la programación basada en Blueprints por su facilidad de uso, aunque en C++ se hace una mejor gestión de la memoria en algunos casos, la implementación de red es más rápida y menos costosa con Blueprints. Podemos ver un ejemplo de un Blueprint en la Fig. 1.

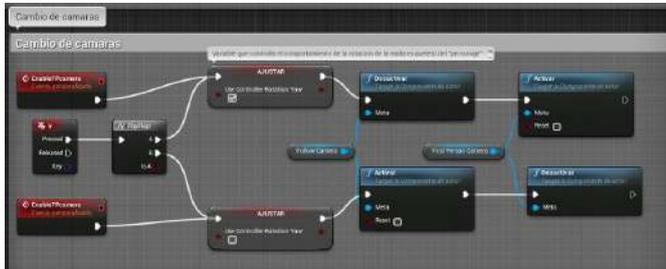


Fig. 1. Ejemplo de Blueprint para salto de personaje.

Para entender cómo se desarrolla con este motor es necesario conocer los elementos que intervienen en las escenas que forman el videojuego.

- A. **Nivel:** En nuestro caso llamamos nivel a una escena en la que los jugadores pueden interactuar con el juego y entre sí con unos límites establecidos
- B. **Malla estática:** Una malla estática es la representación gráfica de un objeto en concreto, lo que normalmente perciben nuestros ojos en la realidad.
- C. **Actor:** mallas estáticas con un script de programación asociado que desarrolla una función, esto permite hacerlos interactivos con el jugador de alguna forma.
- D. **Luces:** las cuales ambientan el nivel y proporcionan luminosidad de carácter general como puede ser la luz del sol o un fuego.
- E. **Actores personajes:** son actores más complejos que se diferencian en que pueden controlarse por un usuario.

Podemos ver un diagrama de estos elementos en la Fig. 2.

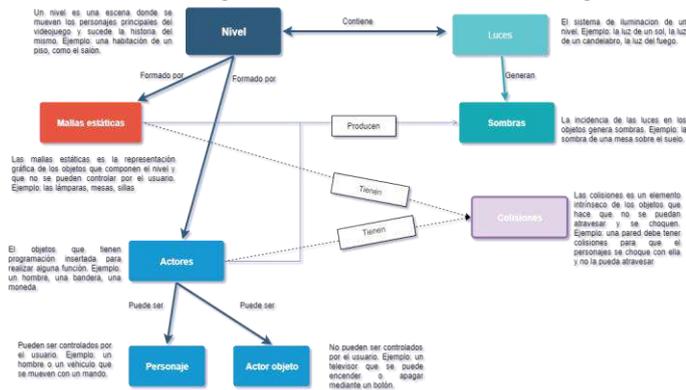


Fig. 2. Elementos de una escena.

## VII. SOPORTE DE RED EN UNREAL ENGINE 4

Una vez tenemos claro el estilo de videojuego que va a desarrollarse es necesario revisar las herramientas con las que contamos para establecer las conexiones a través de la red. Unreal Engine 4 cuenta con diversas maneras de estructurar una partida en línea, las cuales se analizan en profundidad mediante comparativas. Para el videojuego se ha utilizado una modalidad de servidor a la escucha, en la que el servidor de la

partida es a la vez un cliente y puede jugar a la misma vez que hospedar una partida en red. Existen una serie de clases comunes a clientes y servidores que se usan para estructurar la programación y poder replicar la información en la red de forma correcta, aplicando técnicas para controlar el flujo de información. Se ha realizado por tanto un análisis en profundidad de estas clases, a saber, de:

- A. **Game Instance:** es la clase encargada de sustentar el videojuego como entidad y servir de enlace para todas las clases.
- B. **Game Mode:** es la clase encargada de mantener las reglas del juego necesarias en red, encargándose de controlar los eventos para continuar y completar un nivel. Solo accesible desde el servidor.
- C. **Game State:** es una de las clases más importantes en un videojuego en red, mantiene un control en tiempo real de la información compartida del juego en sí, aparte de contener una lista de los jugadores actuales. Accesible desde clientes y servidor.
- D. **Player State:** mantiene un control sobre la información relativa al usuario que forma parte de la red. Usada para la sincronización de clientes mediante el intercambio de variables.
- E. **Player Controller:** es la primera clase que comparten el cliente y el servidor y nos sirve para establecer un enlace entre ambos, de forma que el servidor conoce todos los controladores que hay en el juego sin embargo cada uno de los controladores solo se ve a sí mismo.

Dentro de Unreal Engine 4, la información siempre fluye del Servidor al/los Clientes de forma que todo se centraliza en el equipo servidor, haciendo las comunicaciones seguras impidiendo que los clientes puedan alterar información relevante. Se puede ver un diagrama de la conexión de todas las clases en la Fig. 3. Existen un conjunto de técnicas y modos que permiten que se pueda transmitir la información a nivel de red entre cliente y servidor y de esta forma comunicar las clases citadas, esto es la *replicación*.

Mediante la replicación podemos establecer parámetros como la frecuencia de actualización en red de los actores, el comportamiento ante la saturación del ancho de banda o el porcentaje de descarte de paquetes en algunos casos. De esta forma, por ejemplo, una variable que es replicada se transmite por red sin necesidad de que intervenga otra función que la transmita.

Además de este conjunto de técnicas los clientes se comunican con los servidores mediante llamadas a procedimientos remotos, o por sus siglas en inglés RPC (Remote Procedure Call). Con el uso de todas estas técnicas podremos interconectar usuarios en cualquier parte del mundo y configurar sus conexiones de forma dinámica.

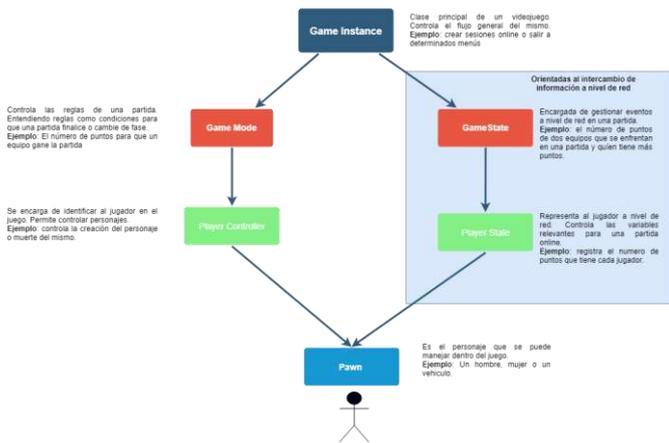


Fig. 3. Relación de clases de red en UE4.

### VIII. TALES OF GRANADA

Dado que es un juego con enfoque didáctico, posee una historia que se basa en el romancero popular granadino y las leyendas sobre la familia de los Abencerrajes. El juego comienza introduciéndonos en la vida de David y María, dos jóvenes granadinos, él encargado de hotel y ella fotógrafa, la cual, tras recibir un importante trabajo para fotografiar los monumentos más destacados de la ciudad, es secuestrada.

Será entonces cuando David y su hermana Alicia deberán trabajar junto con la policía para dar con el paradero de María. Descubrirán que su linaje se remonta al antiguo reino nazarí del siglo XV. Y que están conectados a una serie de intrigas políticas, económicas y amorosas que terminaron con una masacre en pleno palacio de la Alhambra.

Gracias a las notas y a pequeñas pistas que recibirán descubrirán la historia y el poder de los Abencerrajes, una de las pocas familias pertenecientes a una orden musulmana secreta, capaces de canalizar un poder oculto en su sangre y crear fuego mágico, además de otras cualidades singulares.

Con el transcurso de la historia descubrirán que las notas son dejadas por el viajero, un sabio atrapado en el reino mágico del que extraen su poder y con el cual sólo puede comunicarse a través de unas brechas mágicas que interconectan ese reino con la realidad.

Gracias a la espada del sol, capaz de manipular la magia de ese reino, los personajes recorrerán una serie de sitios relevantes para la cultura musulmana y la histórica ciudad de Granada desvelando sus leyendas más antiguas y descubriendo la verdad tras ellas, acercándose cada vez más al paradero de María.

Existen una serie de mecánicas en el juego claramente enfocadas a los dos aspectos principales, docencia y entretenimiento, estas son:

#### A. Mecánica con finalidad educativa, sistema de notas.

Los jugadores tienen que recolectar una serie de ítems cuya representación gráfica es una nota. Mediante estas notas se descubre la historia principal y algunas pistas para resolver los puzzles, este es el método principal de narrativa. Este sistema de notas debe estar siempre accesible al usuario para leerlas y reunir pistas, para ello se ha creado un interfaz denominado "Diario de viaje" el cual posee un registro de las notas cogidas y el objetivo actual que se tiene en un nivel.

#### B. Mecánica con finalidad educativa, sistema de ventanas mágicas.

Existen unas ventanas mágicas de colores que son usadas para guiar a los jugadores y enseñarles cosas sobre la cultura y arquitectura musulmana, los jugadores tan solo deben acercarse para ver esta información como un cartel en la pantalla.

#### C. Mecánica sobre jugabilidad, sistema de pilares.

Los jugadores pueden activar fuentes con magia que les ayudan a resolver puzzles, el servidor (color azul) y el cliente (color rojo) deben alternar los colores de las fuentes para resolver enigmas.

#### D. Mecánica sobre jugabilidad, sistema de estatuas.

Son estatuas que requieren que los jugadores interactúen con ellas con el objetivo de resolver acertijos, pueden ser giradas o movidas pulsando un botón.

### IX. DESARROLLO DEL VIDEOJUEGO

El desarrollo del videojuego ha ido de mayor a menor nivel. De forma general se han programado los comportamientos de las clases de red anteriormente citadas dentro del apartado de soporte de red, interconectando las mismas y estableciendo comunicaciones y mensajes. Podemos ver un ejemplo de función de comunicación mediante un evento multicast en la Fig. 4.

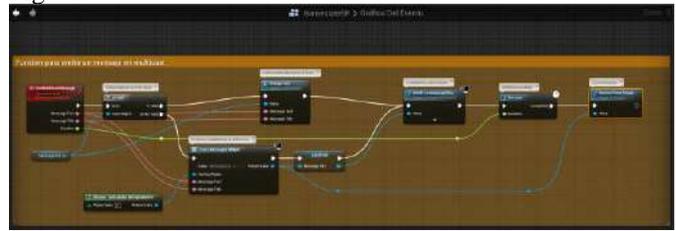


Fig. 4. Ejemplo de función Multicast.

Una vez desarrolladas las clases base del videojuego, se ha programado el comportamiento de los personajes y sus controladores respectivamente.

Más tarde, una vez terminado los personajes y sus funcionalidades vitales para el juego se ha pasado al desarrollo de elementos visuales tales como notas, diseño de mallas, mensajes de alerta, objetivos, menús, sonidos, etc... Viéndose un ejemplo de los mismos en la Fig. 5.



Fig. 5. Personajes (David y María) e interfaz.

Por último se han diseñado los niveles y mallas estáticas necesarias, un ejemplo lo encontramos en la mezquita recreada dentro del primer nivel del videojuego Fig. 6, en la cual podemos ver sus elementos más relevantes (1 – Haram; 2 – Patio exterior /fuentes de abluciones; 5 – Minarete), explicados en el juego mediante notas.

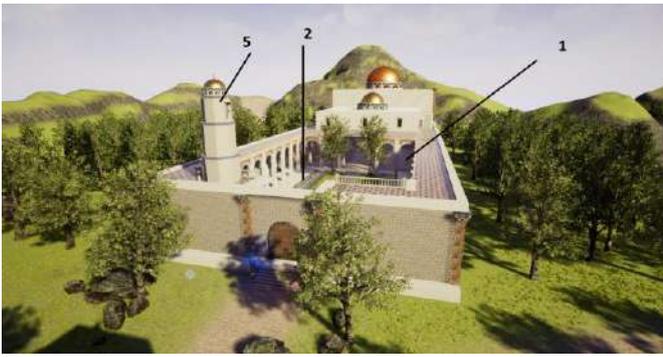


Fig. 6. Mezquita.

X. CARACTERIZACIÓN DEL TRÁFICO DE RED

Uno de los objetivos principales del proyecto es el análisis del tráfico que genera este motor y el videojuego en sí. Por ello, se ha realizado un análisis de los procesos que se llevan a cabo a la hora de establecer y mantener las conexiones entre cliente y servidor.

Se ha analizado de forma práctica el proceso de conexión seguido por un cliente que solicita conectarse al servidor (Fig. 7) y se ha hecho un análisis de los paquetes de red capturados mediante dos herramientas avanzadas para depurar el tráfico de red, Netprofiler y Wireshark, describiendo cada una y cómo se han utilizado para analizar el tráfico que produce el videojuego en todas sus fases.

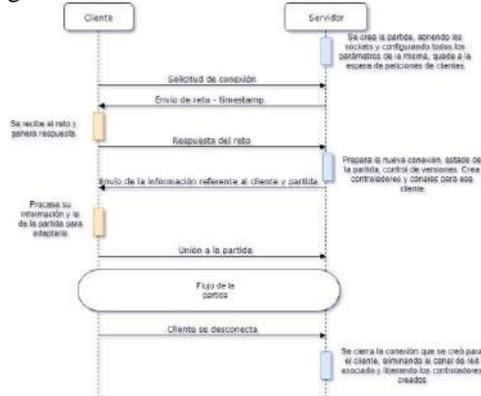


Fig. 7. Conexión Cliente-Servidor.

Podemos ver el entorno de pruebas en el que se ha probado el videojuego en su modalidad de red en Fig. 8, donde nos encontramos a tres clientes y un servidor.

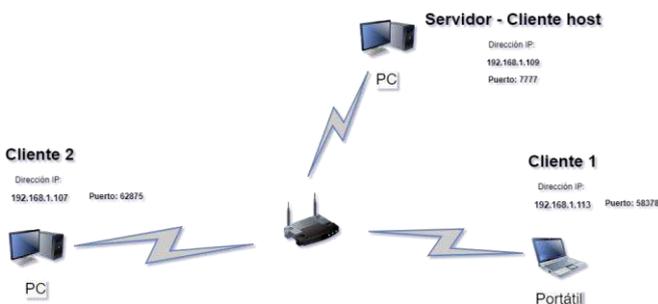


Fig. 8. Entorno de pruebas.

Mediante las herramientas citadas anteriormente se han sacado estadísticas de red en distintas sesiones y equipos clientes y servidores, de forma que podemos desglosar el

consumo de red de manera general y particular para cada clase, función y variable intercambiada por las clases en tiempo real (Fig. 9). Con el uso de Netprofiler, la herramienta para análisis de red propia del motor, podemos extraer los datos que se transmiten a nivel de red en cualquier momento de la conexión y los Bytes transmitidos por todos los elementos por básicos que sean en una comunicación.

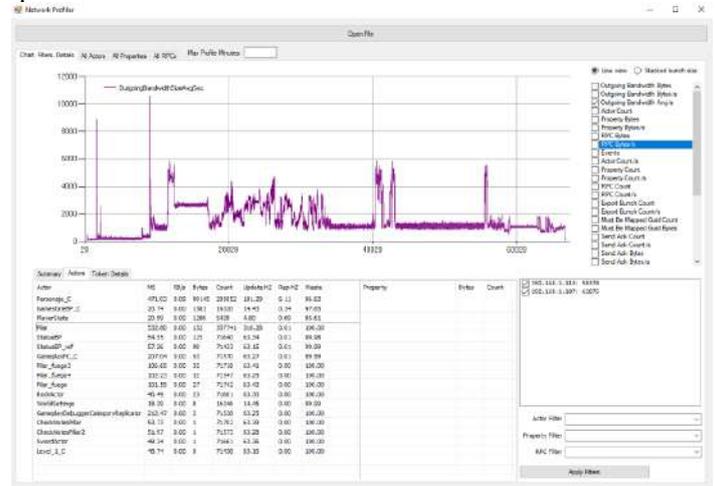


Fig. 9. Tráfico generado por una sesión - Netprofiler.

Además, se han comparado los resultados obtenidos por las dos herramientas, llegándose a los mismos resultados (Fig. 9 - Fig. 10).

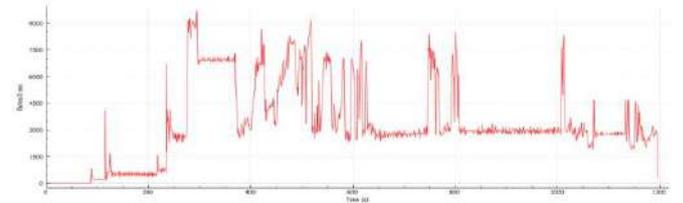


Fig. 10. Tráfico generado por una sesión - Wireshark.

Para finalizar se han obtenido estadísticas de uso y de intercambio de paquetes, además de las longitudes de los mismos y otros datos relevantes como retardos o picos de transmisión de paquetes entre clientes y servidor.

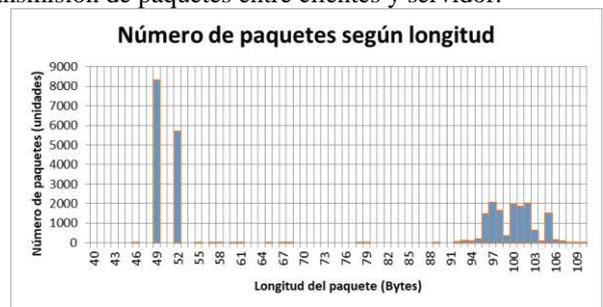


Fig. 11. Número de paquetes según longitud.

XI. CONCLUSIONES A NIVEL DE RED

El ancho de banda final de forma general del videojuego es relativamente bajo, llegando a producir medias de entre 1.5 KB/s y 3 KB/s en diversas pruebas realizadas en clientes y servidor respectivamente. Como es lógico el ancho de banda necesario para el servidor es mucho mayor que para el caso de los clientes puesto que se sigue la filosofía principal de Unreal Engine 4, en la que el servidor se encarga de generar la gran

mayoría del tráfico y el cliente solo manda actualizaciones de estados.

El tamaño medio de los paquetes enviados por el servidor y clientes depende fuertemente del trascurso de la partida y los eventos que sucedan en la misma, sin embargo, tenemos que para el entorno de experimentación ambos, clientes y servidor se mantienen entre los 50 y 80 Bytes, en gran parte influenciado por las RPCs encargadas de transmitir la posición de los jugadores.

Los eventos que más consumen a nivel de red son relativamente pocos, producidos en su mayoría por RPCs, de forma que no llegan a saturar en ningún momento el tráfico o producir un retardo en las transmisiones. Por otro lado, la gran mayoría del tráfico que se produce es debido a la transmisión de las posiciones relativas de cada jugador en el nivel, aunque con una carga relativamente baja de Bytes por segundo se producen de forma continua y periódica.

Se han llevado a cabo optimizaciones de red reduciendo el número de RPCs que se han realizado y en su defecto enviando variables que controlen estados como es el caso del sistema de pilares y su activación.

Otra de las optimizaciones que existen es el envío de referencias de actores en vez de sus correspondientes structs como es el caso del sistema de notas. Con esta optimización se ha conseguido una reducción del ancho de banda que generan los clientes, sin embargo, hemos añadido un coste adicional de computación de lado del servidor.

## XII. METAS ALCANZADAS

### A. Análisis de tráfico de red de un motor de juegos profesional.

En esta meta se ha conseguido estudiar las características del tráfico de red que genera un motor de desarrollo profesional de videojuegos, así como describir las interfaces y herramientas relacionadas con el soporte de red del motor. De esta forma se ofrece dos visiones diferenciadas en alto y bajo nivel. Una de las contribuciones del proyecto es el hecho de que por primera vez dentro de la comunidad hispanohablante de desarrolladores de Unreal Engine 4, se ha hecho un estudio de esta índole, ya que ni siquiera la documentación oficial de Epic Games (la empresa desarrolladora del motor) se encuentra traducida.

### B. Desarrollo de videojuego de carácter didáctico.

Dentro de esta meta se han conseguido implementar una serie de mecánicas que tienen como objetivo narrar de forma entretenida fragmentos de historia y cultura nazarí. No solamente gracias al sistema de notas que cuenta una historia profunda y atrayente, sino que gracias a los portales creados los jugadores aprenderán de arquitectura y religión musulmana de forma dinámica. Dentro de esta meta y gracias a unas primeras pruebas con usuarios podemos afirmar que se ha cumplido con el objetivo

### C. Extra: Localización del videojuego.

Como meta extra se planteó el hecho de ofrecer una traducción a otro idioma y proceder con su localización. Por tanto, el presente videojuego cuenta con localización al inglés.

## AGRADECIMIENTOS

Querría agradecer a mi abuelo Amador, por guiarme hasta tu último suspiro, tus palabras me sirvieron de guía en los peores

momentos, espero que donde quiera que te encuentres puedas verme y estar orgulloso.

A mi pareja Mary por darme fuerzas para alcanzar mis metas, eres la luz que guía mi camino.

A mis padres Elena y Pedro, por darme la vida y apoyarme con todo vuestro ser, anteponiendo las vidas de vuestros hijos a vuestra propia felicidad, sin vosotros nada de esto habría sido posible.

A mi hermano Daniel, porque desde que llegaste nunca más me volví a sentir solo.

A mis tíos Amador y Miguel por enseñarme el verdadero sentido de la palabra familia.

Y sobre todo a mi abuela Fina, tu entereza en la vida es una llama de inspiración para mí.

## REFERENCIAS

- [1] ERA, «Entertainment sales reached £6.3bn in 2016,» [En línea]. Available: <https://eraltd.org/news-events/press-releases/2017/entertainment-sales-reached-63bn-in-2016/>.
- [2] J. Batchelor, «Games industry generated \$108.4bn in revenues in 2017,» 31 enero 2018. [En línea]. Available: <https://www.gamesindustry.biz/articles/2018-01-31-games-industry-generated-usd108-4bn-in-revenues-in-2017>. [Último acceso: 17 junio 2018].
- [3] J. P. Ruiz, «El sector de los videojuegos, el de mayor facturación de la industria digital.» 26 mayo 2017. [En línea]. Available: <https://www.hobbyconsolas.com/noticias/sector-videojuegos-mayor-facturacion-industria-digital-100378>. [Último acceso: 17 junio 2018].
- [4] AEVI, «El videojuego en España,» [En línea]. Available: <http://www.aevi.org.es/la-industria-del-videojuego/en-espana/>. [Último acceso: 17 junio 2018].
- [5] EFECOM, «La industria del videojuego supone el 0,11 % del PIB y emplea 8.790 personas,» 09 01 2019. [En línea]. Available: [https://www.eldiario.es/tecnologia/industria-videojuego-supone-PIB-personas\\_0\\_727577901.html](https://www.eldiario.es/tecnologia/industria-videojuego-supone-PIB-personas_0_727577901.html). [Último acceso: 17 junio 2018].
- [6] J. C. Cortizo Pérez, F. M. Carrero García, B. Monsalve Piqueras, A. Velasco Collado, L. I. Díaz del Dedo y J. Pérez Martín, «Gamificación y docencia: lo que la universidad tiene que aprender de los videojuegos,» agosto 2011. [En línea]. Available: <http://hdl.handle.net/11268/1750>. [Último acceso: 17 junio 2018].
- [7] Ubisoft, «Assassin's Creed,» [En línea]. Available: <https://assassinscreed.ubisoft.com/game/es-es/home/>. [Último acceso: 17 junio 2018].
- [8] R. B. Romana, «Videojuego Libertus "Ruta hacia la libertad",» 09 febrero 2017. [En línea]. Available: <https://beticaromana.org/download/videojuego-libertus-ruta-hacia-la-libertad/>. [Último acceso: 17 junio 2018].
- [9] Unity, «Unity,» [En línea]. Available: <https://unity3d.com/es>. [Último acceso: 17 junio 2018].
- [10] Crytek, «Cryengine,» [En línea]. Available: <https://www.cryengine.com/>. [Último acceso: 17 junio 2018].
- [11] Epic Games, «What is Unreal Engine 4,» [En línea]. Available: <https://www.unrealengine.com/en-US/what-is-unreal-engine-4>. [Último acceso: 17 junio 2018].



**Pedro José Arias Perandrés.** Nacido el 30 de marzo de 1993. Ingeniero en Telecomunicaciones. Apasionado por todo lo relacionado con los videojuegos y especialmente su diseño.

# eGamer.Zone

Autor: Cristian Almohalla Gómez, e-mail: criscricri@correo.ugr.es

Tutor: Juan José Ramos Muñoz, e-mail: jjramos@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—La situación actual del sector de los videojuegos muestra que la cantidad de videojugadores aumenta con el tiempo y que hay un aumento en los ingresos, \$108,000 millones ingresados en 2017.

El objetivo de este proyecto es ofrecer servicios gratuitos a los videojugadores con un modelo comercial realista que se pueda llevar a cabo.

Por un lado, se ha desarrollado el comparador de precios de juegos gratuitos y genera ingresos de los acuerdos de afiliación con tiendas. El comparador se puede ver a través de una página web y a través de una aplicación móvil. Por otro lado, al usuario se le proporciona un servicio de voz sobre Internet Protocol gratuito. Este servicio es usado porque los jugadores necesitan hablar con sus compañeros de equipo mientras juegan en línea. Para el proyecto, este servicio es una forma de darse a conocer.

Como resultado se ha conseguido un conjunto de servicios atractivos para nuestro sector de clientes que forman un modelo de negocio que ya está reportando ingresos que superan el coste de mantenimiento de los servidores y que es fácilmente escalable.

**Palabras clave**—comparador de precios, comunidad, egamers, APP móvil, Servicios web, REST API, egamerzone, firewall, redes, voz sobre IP

## I. INTRODUCCIÓN

ESTE documento es un resumen del trabajo de fin de grado "eGamer.Zone: Diseño e implementación de una plataforma de servicios sobre videojuegos".

En 2017 se ingresaron 108.000 millones de dólares [1], de los cuales 59.000 eran de juegos móviles y 33.000 de PC (Ordenador Personal). En la actualidad está aumentando el número de personas que tienen como hobby los videojuegos. Dentro de este sector de potenciales clientes se quiere plantear servicios que no requieran de grandes inversiones y que permitan plantear productos mínimos viables que se van a publicar, estando sometidos a una continua evaluación para ir mejorando en función del 'feedback' que nos de el cliente.

Se pretende diseñar y desarrollar un servicio concreto que permita desde un coste de inversión bajo empezar a tener beneficios de forma que el proyecto se retro-alimente y evolucione con el capital que produce. Por otro lado, se pretende que sea atractivo para un perfil de cliente que pretende ahorrar y busca ocio en el sector de los videojuegos. Con ese propósito surge la idea de crear un comparador de precios de videojuegos, lo cual será el centro en torno al cual giran el resto de servicios.

En este trabajo, para llevar a cabo este servicio definimos los siguientes objetivos:

- Crear comparador de videojuegos sobre web.

- Trabajar con tecnología API REST (Interfaz de programación de aplicaciones, Transferencia de estado representacional).

## II. ESPECIFICACIONES Y REQUISITOS

EN esta sección se va a definir las necesidades técnicas y funcionales para cumplir los objetivos propuestos de cara al proyecto.

### A. Requisitos Funcionales: Página web

Lugar donde se establece la relación principal con el cliente, en esta el cliente encontrará:

- **Sistema de registro:** el sistema de registro donde el usuario se vincule con nosotros y nos deje la información necesaria para ofrecerle los servicios y contactar con el cuando sea necesario.
- **Sistema de perfil de usuario** y configuración de cuenta: el usuario debe poder en todo momento configurar ciertos parámetros relacionados con su cuenta, como parámetros básicos de privacidad y gestión del registro.
- **Comparador de precios de videojuegos:** esta será la funcionalidad principal y más compleja. El usuario debe poder acceder a un listado de juegos y comparar en una página de producto de un vistazo los precios que ofertan diferentes tiendas.
- **Notificaciones por suscripción:** a través de la web o de la aplicación móvil un usuario podrá activar alertas de precios de cada juego, indicando el precio al que está dispuesto a comprarlo. Al usuario se le notificará por correo y/o a través de la aplicación móvil.
- **Notificaciones por suscripción (no registrados):** usuarios no registrados también podrán activar alertas indicando únicamente su correo. En caso de que dicho usuario se termine registrando utilizando su correo, las alertas anteriormente configuradas se deben asociar a su cuenta.

### B. Requisitos Funcionales: Aplicación móvil

Para llegar a más usuarios se quiere ofrecer una aplicación móvil que de forma sincronizada con la web provea también de la comparativa de precios de videojuegos. Esta debe:

- Presentar de forma sincronizada con la web los juegos del comparador
- Permitir inicio de sesión.
- Permitir suscripción análoga a la de la web para recibir notificaciones de precios.
- Ver los juegos a los que estás suscrito.
- Notificar cuando un juego baja del precio deseado.

### C. Requisitos Funcionales: Servicio de voz sobre IP gratuito

Se pretende ofertar un servidor de voz sobre IP gratuito pensado para jugadores online. Los requisitos este servicio son:

- Tenga una aplicación para PC.
- Consuma poco ancho de banda.
- Permita comunicación grupal sencilla.
- Sea gratuito tanto para el usuario, como para el que lo provee sin ánimo de lucro.
- Tenga un sistema propio por el cual el usuario puede acceder a nuestro servidor.
- Tenga un coste computacional para el servidor bajo.

Se estima que ofertando este servicio los usuarios que lo utilicen vean publicidad de la web y del resto de los servicios ofrecidos.

### D. Requisitos No Funcionales: Servidores públicos en red

Para ofrecer nuestros servicios vamos a necesitar un o varios servidores públicos sobre el cual implementar los diferentes servicios. Estos servidores deberán tener una disponibilidad cercana al 100%.

- Debe tener memoria ( 20Gb), RAM ( 2Gb), computación (1 core dedicado) y ancho de banda suficiente (¿100Mb simétricos) para alojar el servidor de voz.
- Debe tener memoria para almacenar todos los archivos de la web.
- Debe conllevar actualmente un coste de **mantenimiento económico mínimo** hasta que haya un volumen de usuarios considerable que empiece a utilizarlo.

### E. Requisitos No Funcionales: Aplicación servidor web

Es necesario un servicio web que será el eje central. Sobre este se implementarán los servicios principales y el resto de micro-servicios giraran en torno a estos.

- Debe ser gratuito (código abierto).
- Tecnología compatible con el CMS que se vaya a utilizar.

### F. Requisitos No Funcionales: Requisitos de formularios

El sistema de registro y demás formularios debe llevar medidas anti-spam como pueden ser el Google Recaptcha, verificación de e-mail y debe **cumplir con la normativa RGDP**[11], que entró en vigor el 25 de mayo de 2018.

### G. Requisitos No Funcionales: Search Engine Optimization(SEO)

Se debe tener en cuenta los factores de posicionamiento de cara a la realización de la página web para conseguir la mejor visibilidad posible en los motores de búsqueda.

- Utilización de HTML5 y etiquetas semánticas.
- Inserción de datos estructurados.
- Inserción de tarjetas enriquecidas.
- Versión de páginas AMP(Acelerate Mobile Pages) para productos.
- No presentar errores 40X y 50X para páginas anteriormente indexadas o actualmente indexables.
- Generar XML Sitemap para facilitar la indexación a los motores de búsqueda.
- Generar robots.txt para evitar que motores de búsqueda indexen recursos no deseados.

### H. Requisitos No Funcionales: Interfaz de programación de aplicaciones, Transferencia de estado representacional (API REST)

Para que los usuarios de la web y los productos del comparador estén sincronizados con la aplicación móvil se debe crear una API REST que consulte la misma base de datos que la web y ofrezca los recursos necesarios para que la aplicación móvil implemente las diferentes funcionalidades.

### I. Requisitos No Funcionales: Adaptación de la web a móvil

La web se debe programar de forma que se adapte a las pantallas de los móviles y tablets, esto se consigue a través de Bootstrap [4].

## III. PLANIFICACIÓN Y COSTES

Para la planificación del trabajo se hizo un diagrama de Gantt por fases de desarrollo que se puede ver en la memoria del proyecto.

Respecto a los costes, este proyecto conlleva gastos de mantenimiento de dominio y servidores que se pagan con suscripciones de usuario. Por otro lado, se tiene los ingresos obtenido a partir de los contratos de afiliación con las tiendas de videojuegos.

En este momento el proyecto se mantiene a si mismo, no generando pequeños ingresos, una vez descontado el mantenimiento. No obstante, para estimar el coste real del proyecto debemos tener en cuenta el salario que habría sido necesario para pagar la fuerza de trabajo, que se puede estimar en 1.400 euros mensuales incluyendo impuestos, durante 8 meses a media jornada. Lo que hace un coste total de 11.200 euros.

## IV. ANÁLISIS

**S**E han estudiado las diferentes plataformas que puedan ofrecer servicios similares al propuesto. También se han revisado las diferentes tecnologías que permitan ofrecer los servicios de acuerdo a las especificaciones.

### A. Estado del arte

De cara a un proyecto de emprendimiento el principal error que se suele cometer es pensar que nuestra idea es única. Y cómo es lo habitual podemos encontrar plataformas que ofrecen servicios similares al expuesto:

- Godkeys [26].
- Gamerastreator [28].
- ClaveCD [6]: este se acerca más a la idea que se tenía de un comparador de precios. Como puntos a mejorar: no es una web atractiva y su algoritmo de relación de juegos falla con los juegos que tienen nombres similares, mostrándolo como un único producto.

### B. Alojamiento de servidores

Los servicios se van a organizar en varios micro-servidores en los que se van a estar distribuidos. Se han elegido esto micro-servidores por su precio económico en relación a las tarifas de servidores con especificaciones poco superiores. La empresa elegida fue ArubaCloud[30].

Se han probado diversos proveedores españoles (con centro de datos en España) antes de decidirse por el proveedor actual. Los proveedores españoles además de ser mucho más caros, daban problemáticas de cara a poder mantener un servicio con una disponibilidad cercana al 100%. Los proveedores probados y descartados:

- TuTiendaHosting [32].
- Hostinger [24].
- Cubenode [10].
- Hostigal [23].

En la última etapa del proyecto se consiguió un alquiler compartido de un servidor Cloud con recursos dedicados con un tercero en la empresa Hostalia [22], la cual si tiene alojamiento en España.

### C. Servicio voz sobre IP

Las alternativas para el servicio del voz IP planteadas:

- Course Client.
- RaidCall [29].
- Discord [9].
- Team Speak 3 (TS3)[31].

Se ha elegido TeamSpeak 3 por la flexibilidad en la configuración que ofrece y la capacidad de dar nuestro servidor por parte del servicio sin coste.

### D. Content Management System (CMS)

Se ha decidido usar un CMS, lo cual nos permite partir de un sistema de gestión básico sobre el cual desarrollar el resto de funcionalidades concretas. Se planteó la posibilidad de usar Drupal [7], Joomla [8] o Wordpress [25].

Se decidió utilizar Wordpress por:

- Ser el más usado y por tanto, tener una comunidad de desarrolladores mucho mayor.
- Sistema de gestión sencillo de base y personalizable a todos los niveles.
- Facilidad para aplicar actualizaciones sin incompatibilidad entre versiones.

### E. Aplicación Móvil

Para desarrollar la accesibilidad del comparador a usuarios en móviles se plantean tres posibilidades:

- Crear aplicación nativas que sean muestren la web en versión web.
- Crear una aplicación mediante un framework que permita del desarrollo de aplicaciones multi-plataforma (Android e IOS) y crear una aplicación que muestre los productos del comparador de precios a través de una API que sincronice la página web y la aplicación.
- Adaptación de la web a móvil mediante Bootstrap y utilización de la tecnología Prograssive Web aplicación (PWA).

Se decidió llevar a cabo las dos últimas opciones. Por un lado adaptar la web completamente a móviles y por otro lado crear una aplicación a través de un framework multi-plataforma.

## V. DISEÑO

EN esta sección se va ver el diseño planteado para la arquitectura global de red, las interconexiones entre los diferentes servicios, así como la configuración específica de cada subsistema, así como otros aspectos generales.

### A. Seguridad

La imagen del proveedor de servicios de Centos 7 x64 montada sobre VMWare en el servidor no dispone de un servicio **firewall** instalado de serie. Se ha instalado firwallld en todos los servidores, bloqueando todos los puertos y tipos de tráfico que no se utilizan.

El cloud server además dispone de un **firewall físico** que protege al servidor frente ataque DoS (denegación de servicio) entre otros.

También se ha previsto de un **firewall de aplicación** a la página web, el cual revisa y puede filtrar el tráfico a esta si detecta actividades maliciosas.

Se ha llevado a cabo de un **anti-malware** para proteger el servidor de software malicioso que pueda introducirse a través de la web, el servicio de correo, etc.

Por último se ha llevado a cabo una **configuración extendida** del servicio **SSH** (Secure Shell):

- Cambio de puerto 22 a 2222
- Usuario alternativo a root para el acceso con capacidad de super usuario.
- Bloqueo de inicio de sesión remoto con el usuario root.
- Conexión al servicio a través de un par de claves asimétricas.
- Eliminación de la posibilidad de iniciar sesión con contraseña.
- Límite de intentos de conexión y sesiones simultáneas.
- Uso Fail2Ban: firewall de aplicación para SSH.

### B. Arquitectura global

La arquitectura ha evolucionado durante la realización del proyecto, adaptándose a los requisitos de los diferentes servicios o a la posibilidades económicas fruto de nuevos acuerdos. El diseño actualmente vigente se puede ver en la sección análoga de la memoria del proyecto.

### C. Subsistema voz sobre IP

Este servicio permite la configuración avanzada de permisos de usuarios y requiere de una configuración de registros y servicios DNS (Domain Name Server) especial si queremos acceder al servidor con un nombre de dominio. Como contra es un sistema muy complejo, pero gracias a esto nos permite añadir nuevas formas de funcionar dentro del servidor o, por la parte de la configuración DNS, acceder al servidor de Team Speak 3 usando el dominio base 'egamer.zone' sin que ese dominio este apuntando a la IP donde está el servidor Team Speak 3. Por tanto se ha llevado a cabo el diseño de permisos y el diseño de red de cara a este servicio.

Además se ha diseñado un sistema de "backups" y "anti-crash": para hacer copias de seguridad periódicamente y reaccionar automáticamente ante caídas del servicio, optimizando al máximo la disponibilidad de este.

#### D. Subsistema Web

La base central de nuestro modelo de negocio va a estar en la página web.

Para la página web se ha elegido Wordpress como CMS, siguiendo su filosofía de programación se debe crear un tema hijo que herede las plantillas de un tema previamente elegido. Sobre dicho tema hijo se implementarán las modificaciones estéticas deseadas y las funcionalidades más simples. Las funcionalidades complejas se organizarán en lo que llaman plugins, en su mayoría desarrollados por terceros (open source).

Para el comparador de precios crearemos un plugin específico que trabajará como extensión del plugin de e-commerce más utilizado en Wordpress, Woocommerce[33].

La relación entre el CMS, los plugins y el tema lo podemos ver más claramente en una figura en la sección análoga de la memoria del proyecto.

A grandes rasgos, la web tendrá los siguientes apartados:

- Se accederá desde los dominios: egamer.zone y www.egamer.zone
- Implementará conexión segura mediante el protocolo HTTPS
- Permitirá a los usuarios registrarse, validando el correo electrónico.
- Implementará el servicio principal (comparador de precios de videojuegos) mediante un plugin.
- Permitirá suscripción a juegos (activación de alertas de precios) por parte del usuario.
- Se va a optimizar para mejorar el posicionamiento y el rendimiento, consiguiendo así tráfico orgánico y la reducción de la tasa de rebote.
  - Optimización de imágenes.
  - Estructura semántica del código HTML5.
  - Minificación de recursos: html, css, js.
  - Implementación AMP.
  - Creación de sitemaps.
  - Localizar los recursos js y css al final de la página para que no bloqueen su carga.
- Sincronización con una API Rest que será consultada a través de una aplicación móvil que se ofrecerá también a los usuarios.
- Dispondrá de un sistema de gestión de usuarios: registro, confirmación de correo, acceso y gestión de cuenta y perfil.

#### E. Comparador de precios

La dificultad principal para realizar el comparador está en la relación de los productos iguales en diferentes tiendas. Para ello se ha llevado a cabo un formateo de la información de cada juego, lo que facilita el posterior trabajo con esta.

Luego se han planteado varios métodos para llevar a cabo esta relación:

- **Método automático:** teniendo en cuenta el nombre y la plataforma del juego, relacionará dichos juegos si la coincidencia es total, es decir, si pertenecen a la misma plataforma y tienen exactamente el mismo nombre.
- **Método semi-automático** (usado cuando el inventario de juegos de la tienda a relacionar es superior al inventario de nuestro comparador): se encontrará el juego más

probable para cada juego del inventario del comparador, teniendo en cuenta que el nombre y la plataforma, y se le pedirá al usuario administrador que confirme que dicha asociación es correcta con un campo check.

- **Método semi-automático (inverso)** (usado cuando el inventario de juegos de la tienda a relacionar es inferior al inventario de nuestro comparador): es la alternativa al método semi-automático, funciona igual. La diferencia es que coge como punto de referencia el inventario de la tienda que estamos relacionando y busca el más probable en el inventario del comparador.
- **Método manual** (usado cuando el inventario de juegos de la tienda a relacionar es superior al inventario de nuestro comparador): para cada juego del inventario del comparador habrá que indicar la URL (Uniform Resource Locator) y la ID (identificador) del juego de la tienda específica mediante un formulario.
- **Método manual (inverso)** (usado cuando el inventario de juegos de la tienda a relacionar es inferior al inventario de nuestro comparador): con la misma filosofía que el anterior método inverso, este método es igual que el manual pero se parte del inventario de juegos de la tienda específica y hay que indicar con que juego del inventario del comparador hay que relacionarlo.

#### F. Subsistema Aplicación móvil

Se va a crear un menú tipo hamburguer muy común en las aplicaciones. Desde este menú se puede acceder a las características que la aplicación poseerá en un primer momento:

- inicio de sesión.
- buscador de juegos.
- alertas de precios.

El contenido del menú es dinámico en función de si el usuario ha iniciado o no sesión.

Esta aplicación será programada con IONIC, es una empresa pionera en el desarrollo de soluciones tecnológicas móviles, especializada en desarrollar productos móviles para la industria bancaria y financiera latino americana. La aplicación se conectará a una API Rest programada en mi web de la cual obtengo la información de los productos en JSON.

## VI. IMPLEMENTACIÓN

SE ha llevado a cabo la implementación de todo el diseño expuesto, y se explica paso a paso detalladamente como se ha hecho en la memoria del proyecto.

## VII. PRUEBAS

SE han realizado varias pruebas para probar cuantificar la calidad de los resultados obtenidos. Las gráficas de resultados se encuentran en la memoria del proyecto, en este resumen se describe brevemente en que consisten dichas pruebas.

#### A. Disponibilidad

Nuestro proveedor de servicio nos garantiza una disponibilidad del 99%, lo que quiere decir que, como mínimo,

el 99% del tiempo, nuestro servidor estará en completo funcionamiento.

De cara al servidor de voz sobre IP nosotros hacemos un mantenimiento adicional de cara a la realización de Backups, de manera que para este servicio la disponibilidad es inferior a ese 99%, y es por eso que vamos a calcularla.

El backup se hace una vez al mes, vamos a partir de que cada mes tiene 30 días y que el backup mantiene el servicio desconectado durante 30 segundos. Por otro lado, aunque no se da con esa frecuencia, vamos a asumir que el servicio se interrumpe por error de la aplicación servidor una vez al mes. Con el script "anti-crash" el servicio se levantaría a los 1-59 segundos, ya que la comprobación se hace cada minuto y dependería del momento del error. Es decir, tenemos dos errores cada 30 días de una media de duración de 30 segundos.

$$\text{Disponibilidad}(D) = \frac{MTBF}{MTBF+MTTR}$$

Dónde MTBF es el tiempo medio entre fallos y MTTR es el tiempo medio para reparar la falla (30 segundos en nuestro caso).

$$MTBF = \frac{\text{tiempodeoperacion}-\text{tiempodefallas}}{\text{n}^{\circ}\text{defallas}} = \frac{30\text{días} \times 24\text{horas} \times 3600\text{segundos} - 60\text{segundos}}{2} = 1295970s$$

De forma que nuestra Disponibilidad total (DT) a partir de nuestra Disponibilidad del proveedor (DP) es:

$$DT = DP \times D = DP \times \frac{MTBF}{MTBF+MTTR} = 0.99 \times \frac{1295970}{1295970+30} = 0.98998 \quad DT = 98.998\%$$

### B. Prueba Rendimiento

Para mejorar el rendimiento de nuestra Web Google tiene una herramienta llamada PageSpeed Insights(PSI)[16][19], esta proporciona datos sobre el rendimiento real de las páginas tanto en dispositivos móviles como en ordenadores, y da consejos para mejorarlas.

### C. Prueba Datos Estructurados

Los datos estructurados se forman a partir de etiquetas semánticas del código HTML5 y a partir de Tarjetas enriquecidas, que son objetos en Json que describen el contenido de la página. La estructura de estos datos tiene un formato "estándar" que viene dado por Schema.org[21]. No se le denomina "estandar" pero es el reconocido por los motores de búsqueda entre otras aplicaciones de Google, Microsoft, Pinterest, Yandex,...

### D. Pruebas Accelerated Mobile Pages

De forma similar al anterior punto, la estructura y la forma de programación de Accelerated Mobile Pages (AMP) viene dada por un "estandar" no reconocido como tal, ya que la tecnología AMP surge de Accelerated Mobile Pages Project[18], en este caso al ser impulsado por Google se ha terminado imponiendo y a día de hoy muchas páginas importantes lo implementan. Por poner un ejemplo a nivel nacional, los principales periódicos implementan AMP: El Mundo, El País, El ABC, Público,...

Google también nos da una herramienta[17] para comprobar la validez de la estructura, scripts y etiquetas usadas en la implementación.

### E. Prueba Progressive Web App

Para probar que cumplimos los requisitos para decir que nuestra web es una Progressive Web App hemos utilizado la documentación de Google[20]. Aquí explica como cumplir los requisitos específicos y nos proporciona una herramienta para evaluar nuestra web como PWA.

### F. Análisis de tráfico de Team Speak 3

En un primer lugar, es necesario analizar las consultas DNS[13] que realiza para a partir de estas poder diseñar nuestra red de servicios conforme a nombres de dominio. El cliente de Team Speak 3 hace una consulta DNS tipo SRV[14] para obtener el dominio que apunta a un servicio DNS propio que implementa el servidor de TS3. Luego consulta la IP del servidor del servicio DNS propio con una consulta tipo A. Finalmente hace una consulta a su servicio DNS propio para obtener la IP y el puerto del servidor DNS asociado al subdominio.

También se han llevado a cabo pruebas de consumo de ancho de banda del usuario en varias situaciones diferentes, obteniendo gráficas de este a partir del tráfico capturado.

## VIII. SITUACIÓN DEL MERCADO

**E**L contexto del proyecto es favorable y prometedor, año tras año se suceden las noticias en las que la venta de videojuegos no hace si no mejorar.

Periódicos de tirada nacional como son El Mundo[12], El País[5], RTVE[15], 20 minutos[2], ...[3] venían anunciando el crecimiento de un sector que informe de Entertainment and Media Outlook 2017-2021[27], confirmando la tendencia, prevé que se gastará un 3,5% más en este sector para 2021.

El informe no solo apunta al aumento de facturación, también hace hincapié en las transformaciones aceleradas que se dan en este sector. La digitalización rompen barreras con el consumidor facilitando la adquisición de productos, pero a su vez, fragmenta el mercado y hacen necesario el cambio de modelo de negocio a la vez que abre puertas a nuevos modelos.

En definitiva, el sector esta en evolución y expansión, lo cual plantea un terreno favorable de cara llevar el negocio planteado adelante.

En la memoria del proyecto se puede encontrar un análisis del mercado mucho más extendido y también un análisis del cliente potencial.

## IX. CONCLUSIONES

**S**E puede decir que se han logrado todos lo objetivos planteados para este proyecto, matizando la idea inicial que se tenía en función del feedback recibido por los usuarios y las hipótesis que se han ido trabajando hasta tener el resultado que actualmente se presenta:

- Desarrollar un interfaz web personalizaba básica, intuitiva y bastante atractiva tanto en la página web como en la aplicación.
- Obtener el contrato de afiliación con ocho tiendas de videojuegos.
- Extraer la información de las diferentes tiendas con el método específico que nos ofrezca cada una o mediante una araña que analice su web (scraping).

- Relacionar los juegos de las tiendas de forma que la probabilidad de error (de que los juegos relacionados no sean exactamente el mismo) sea prácticamente cero y que sea lo más automático posible, generando varios métodos de relación de productos.
- Se han optimizado los recursos de la web y se ha llevado a cabo una programación estructurada y con una filosofía correcta que permite la actualización de los módulos de terceros modificados sin perder dichas modificaciones.
- Desarrollar una aplicación móvil que no requiera un mantenimiento excesivo e implemente el comparador con notificaciones de forma totalmente sincronizada con la página web.
- Se ha trabajado con trabajado con tecnología API REST para llevar a cabo la sincronización de la web y la aplicación.
- Prestar un servicio voz sobre IP gratuito con el que empezar a darse a conocer.
- Generar una pequeña comunidad de jugadores que masificada puede llegar a ser un negocio muy rentable.

Cabe destacar que todos los productos desarrollados están actualmente publicados y siendo usados por usuarios a excepción de la aplicación móvil, y que como se puede deducir de la descripción del proyecto, se han conseguido muchos más objetivos de los listados anteriormente.

#### AGRADECIMIENTOS

Agradecimientos a mi familia, si su apoyo no habría podido llevar a cabo mis estudios.

Agradecimientos a mis compañeros y amigos, esas personas con los he compartido trinchera cuatro años, con las que he aprovechado, perdido, disfrutado, sufrido,... el tiempo. Sin el trabajo en equipo ninguno habríamos llegado al final.

Recordar a los compañeros que no pudieron terminar la carrera, por verse superados por la exigencia de esta o por motivos económicos.

Agradecimiento a la universidad, pública y para todos, y a los profesores que dan clase a sus alumnos con entusiasmo y dedicación.

Por último, agradecer a todos los que han apoyado económicamente con una pequeña cantidad para pagar el mantenimiento de los servidores sobre los que se implementan todos los servicios de este proyecto.

#### REFERENCIAS

- [1] 20MINUTOS.ES. Los videojuegos supusieron un negocio de 108.000 millones de dólares en 2017. 20 minutos: <https://www.20minutos.es/videojuegos/noticia/videojuegos-2017-negocio-mundo-mas-100000-millones-dolares-3250000/>. Última visita: agosto 2018, 2018.
- [2] AGENCIAS. El gasto de los españoles en videojuegos ha crecido un 17año. 20 minutos: <http://www.rtve.es/noticias/20170608/industria-del-videojuego-espana-facturo-1163-millones-2016-74-mas/1561938.shtml>. Última visita: agosto 2018, 2018.
- [3] Álvaro Alonso, Antonio M. Martín, q256, and jiXo. Enlaces a varias noticias que auxiliares que revelan el avance del sector de los videojuegos. El otro lado: [https://www.elotrolado.net/noticia\\_aumenta-la-venta-de-videojuegos-en-medio-de-la-crisis\\_14983](https://www.elotrolado.net/noticia_aumenta-la-venta-de-videojuegos-en-medio-de-la-crisis_14983), Vida Extra: <https://www.vidaextra.com/industria/la-industria-del-videojuego-bate-records-de-ventas-en-2008>, El otro lado: [https://www.elotrolado.net/noticia\\_la-crisis-economica-afecta-a-la-industria-de-videojuegos-en-espana\\_15912](https://www.elotrolado.net/noticia_la-crisis-economica-afecta-a-la-industria-de-videojuegos-en-espana_15912), El Diario <https://www.eldiario.es/juegoreviews/noticias/> industria-videojuego-Espana-crecido-ano-pasado\_0\_369563649.html, El independiente: <https://www.elindependiente.com/economia/2017/11/20/videojuegos-ventas-espana/>. Última visita: agosto 2018, 2008 - 2017.
- [4] Fakhri Alsadi. Bootstrap. <https://getbootstrap.com/>. Última visita: agosto 2018.
- [5] Susana Carrizosa. Videojuegos para no salir de casa. El País: [https://elpais.com/economia/2016/04/08/actualidad/1460109081\\_254128.html](https://elpais.com/economia/2016/04/08/actualidad/1460109081_254128.html). Última visita: agosto 2018, 2015.
- [6] claveCD. Clavecd. <https://www.clavecd.es/>. Última visita: marzo 2018.
- [7] Community and partners. Drupal. <https://www.drupal.org/>. Última visita: marzo 2018.
- [8] Community and partners. Joomla!. <https://www.joomla.org/>. Última visita: marzo 2018.
- [9] Gamers Community. Discord. <https://discordapp.com/>. Última visita: junio 2018.
- [10] Cubenode. Cubenode. <https://cubenode.com>. Última visita: septiembre 2018.
- [11] Agencia Española de protección de datos. Responsabilidad proactiva. <https://www.aepd.es/reglamento/cumplimiento/principio-responsabilidad-proactiva.html>. Última visita: agosto 2018, 2018.
- [12] Jesús Domínguez. Los españoles gastaron 318 millones de euros en videojuegos en 2015. El mundo: <http://www.elmundo.es/economia/2016/03/30/56fbedb5e2704e11248b45c6.html>. Última visita: agosto 2018, 2015.
- [13] RFC Editor. Domain names - concepts and facilities. <https://www.ietf.org/rfc/rfc1034.txt>. Última visita: agosto 2018, 1987.
- [14] RFC Editor. A dns rr for specifying the location of services (dns srv). <https://www.ietf.org/rfc/rfc2782.txt>. Última visita: agosto 2018, 2000.
- [15] RTVE.es / EFECOM. La industria del videojuego en España facturó 1.163 millones en 2016, un 7,4 RTVE: <http://www.rtve.es/noticias/20170608/industria-del-videojuego-espana-facturo-1163-millones-2016-74-mas/1561938.shtml>. Última visita: agosto 2018, 2016.
- [16] Google. Pagespeed insights. <https://developers.google.com/speed/pagespeed/insights/?hl=es>. Última visita: agosto 2018.
- [17] Google. Prueba de amp. <https://www.google.com/chrome/>. Última visita: agosto 2018.
- [18] Google. Accelerated mobile pages project. <https://www.ampproject.org/es/>. Última visita: agosto 2018, 2015.
- [19] Google. Pagespeed insights. <https://developers.google.com/speed/docs/insights/about?hl=es-419>. Última visita: agosto 2018, 2018.
- [20] Google. Progressive web app checklist. <https://developers.google.com/web/progressive-web-apps/checklist>. Última visita: agosto 2018, 2018.
- [21] Community Group and Steering Group. Schema.org. <https://schema.org/>. Última visita: agosto 2018.
- [22] Hostalia. Hostalia. <https://www.hostalia.com/>. Última visita: agosto 2018.
- [23] Hostigal. Hostigal. <https://www.hostigal.com/cloud-hosting/>. Última visita: septiembre 2018.
- [24] Hostinguer. Hostinguer. <https://www.hostinger.es>. Última visita: septiembre 2018.
- [25] Matt Mullenweg and Community. Wordpress. <https://wordpress.org/>. Última visita: marzo 2018.
- [26] overthings. Goodkeys. <https://goodkeys.com/es>. Última visita: marzo 2018.
- [27] PCW. Entertainment and media outlook 2017-2021 España. <https://www.pwc.es/es/publicaciones/entretenimiento-y-medios/assets/gemo-espana-2017-2021.pdf>. Última visita: agosto 2018, 2018.
- [28] Enrique Somolinos Pérez and Máximo Somolinos Pérez. Gamas-treator. <http://www.gameras-treator.com/>. Última visita: marzo 2018.
- [29] RaidCall. Raidcall. <http://www.raidcall.es/>. Última visita: junio 2018.
- [30] Aruba S.p.A. Arubacloud. <https://www.arubacloud.es/>. Última visita: agosto 2018.
- [31] Team Speak. Team speak 3. <https://www.teamSpeak.com/en/>. Última visita: agosto 2018.
- [32] TuTiendaHosting. Tutiendahosting. <https://tutiendahosting.com/>. Última visita: septiembre 2018.
- [33] Woocommerce and Community. Woocommerce. <https://woocommerce.com/>. Última visita: julio 2018.

# Juego de Estrategia Multiplataforma con Almacenamiento en la Nube: Cross General

Autor: José Carreto Morales, e-mail: jocamo@correo.ugr.es

Tutor: Juan José Ramos Muñoz; e-mail: jjramos@ugr.es

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones

Universidad de Granada

**Resumen**—La industria del videojuego ingresó 108'9 millones de dólares en el último año. A día de hoy, los videojuegos son considerados uno de los grandes mercados del entretenimiento a nivel global junto a las industrias tradicionales del cine o la música.

Gran parte de este éxito viene motivado por su presencia en prácticamente cualquier plataforma. No obstante, esto ha generado un panorama de gran fragmentación. Resultan verdaderamente escasos los videojuegos inter-plataforma que permitan el juego cruzado entre usuarios de distintos dispositivos.

En el presente Proyecto se diseña y desarrolla un juego multijugador que se adapta a distintas plataformas (Android y PC). El videojuego ofrece conectividad y juego inter-plataforma, así como un sistema de almacenamiento de progreso en la nube. Estas características no solo permiten al videojuego llegar a más público, si no que lo transforman en una herramienta que, al fin, es capaz de unir a comunidades y jugadores.

**Palabras clave**— Android, CrossPlay, Guardado, Inter-plataforma, Juego Cruzado, Multiplataforma, Nube, RakNet, RPC, Unity, Videojuego.

## I. INTRODUCCIÓN

En este proyecto se ha realizado el diseño, desarrollo e implementación de un videojuego multijugador inter-plataforma con guardado en la nube. Durante el trabajo efectuado, se ha realizado un estudio completo de los diferentes campos a tener en cuenta en un desarrollo de estas características, como los tipos de videojuegos existentes, el mercado de dispositivos actual o las herramientas disponibles para el desarrollo.

El resultado del proyecto es un videojuego de estrategia basada en turnos, multijugador y disponible en dos de las plataformas más usadas en el mundo: Android y PC (Windows). Las versiones de estas plataformas han sido desarrolladas para permitir la interacción entre ellas. Esto quiere decir que un jugador con un Smartphone Android, puede participar en la misma partida que un jugador con un PC Windows, algo que en la actualidad pocos videojuegos permiten.

Así mismo, el videojuego cuenta con un sistema de almacenamiento en la nube para permitir salvar el progreso de la partida de forma remota. Este progreso o estado de partida es almacenado en ficheros creados específicamente por el videojuego. Estos archivos son totalmente compatibles

con ambas plataformas, es decir, un jugador puede guardar su partida en curso en un Smartphone y continuarla en un PC más tarde. La disponibilidad total de los archivos de guardado, gracias a que se encuentran en un servidor conectado a Internet, sirve para reforzar aun más la cohesión entre las distintas plataformas.

El trabajo se ha enfocado hacia la simulación de un producto completo. No solo se ha desarrollado un videojuego sin más, si no que se ha puesto especial detalle en elaborar y/o estudiar los elementos complementarios del mismo como: el apartado de red, las plataformas de venta y marketing o el formato de distribución.

## II. REVISIÓN DEL ESTADO DEL ARTE

En este apartado, tras la introducción, se busca realizar una revisión general del estado del arte asociado a este proyecto, además, servirá para explicar y aclarar ciertos conceptos relativos al desarrollo del mismo. Los puntos claves a revisar son: el género de videojuego, los motores de desarrollo, el concepto de Crossplay y Cloudsaving, así como las diferentes plataformas y S.O. disponibles.

### A. Multiplataforma y Crossplay

Los conceptos de *multiplataforma* y *Crossplay* son a menudo confundidos y esto puede dar lugar a grandes malentendidos.

Cuando un juego se desarrolla se hace con vistas a su publicación en un dispositivo o S.O. concreto, a estos dispositivos y S.O. se les conoce como plataformas objetivo. Un videojuego multiplataforma será aquel que cuente con versiones para varias de dichas plataformas.

En cuanto al concepto de *Crossplay*, o como también se le conoce, *Crossplatform*, este hace referencia a la posibilidad de que un juego de una plataforma pueda interactuar con otra copia del mismo, pero de una plataforma distinta. Es decir, un usuario de por ejemplo PC, podría estar jugando a un videojuego con un usuario que hace uso de una consola como la *Playstation 4*. Por esta razón, para que un juego tenga *Crossplay*, primero, deberá ser *multiplataforma*. Se puede concluir que no todos los juegos *multiplataforma* disponen de *Crossplay*, pero sin embargo, sí que todos los juegos que disponen de *Crossplay* deben ser obligatoriamente *multiplataforma*.

### B. Actualidad del concepto de Crossplay

Los títulos con esta característica son realmente escasos, no obstante, los que se encuentran actualmente disponibles, gozan de gran fama y público. Conviene destacar el caso del videojuego *Fortnite*.

*Fortnite* es el caso de mayor éxito de un videojuego que cuenta con Crossplay. Su desarrollo se ha realizado de forma progresiva durante estos últimos años y actualmente se encuentra en el Top 5 de los videojuegos con mas jugadores.

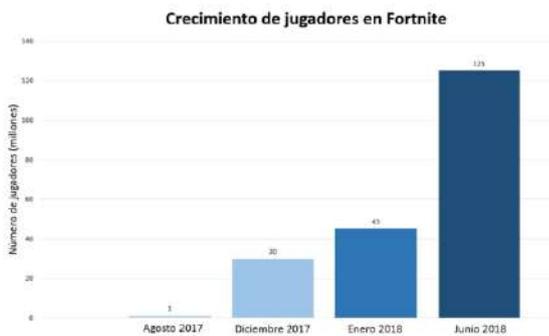


Fig. 1. Número de jugadores (en millones) de *Fortnite* a nivel mundial. Datos extraídos de: [1].

En Agosto de 2017 contaba con 1 millón de usuarios y desde entonces su número no ha dejado de crecer. El aumento de jugadores hasta el día de hoy puede verse reflejado en los datos de la Fig. 1. Su gran popularidad se debe a su carácter gratuito, cualquiera que lo desee debe simplemente descargarse su instalador desde la página oficial y tras su descarga e instalación estará en disposición de jugar en solitario o en equipos de hasta 4 jugadores. Su carácter de juego rápido y altamente competitivo, unido a su estilo nuevo y desenfadado lo han convertido en un auténtico fenómeno de masas.

*Fortnite* puede presumir de ser el único videojuego que está presente en casi la totalidad de las plataformas actuales de juegos: PC, Playstation 4, Xbox One, Nintendo Switch y móviles/tablets con iOS y Android.

### C. Almacenamiento en Nube (Cloud Storage)

El almacenamiento en la Nube, o Cloud Storage como se le conoce al término en inglés, es un modelo de servicios en el que los datos de un sistema o arquitectura son almacenados, administrados y asegurados (mediante copias de respaldo) de forma remota. Generalmente los servidores o equipos remotos que poseen estos datos están en línea, bien de manera pública y accesible vía Internet o dentro de redes corporativas privadas. [2]

Actualmente hay una gran variedad de plataformas de *Cloud Storage* disponibles tanto para empresas como para particulares. Hay soluciones generalistas como *Dropbox* o *Mega*, pero también hay plataformas de almacenamiento mas especializadas como puede ser el caso de *Git* o *Github*, ambas enfocadas a la programación. En este proyecto en concreto, el guardado remoto, o en la nube, permite cohesionar el uso de varias plataformas, ya que mediante su utilización se podrá salvar el progreso logrado en una

plataforma para poder continuar con el mismo en una plataforma distinta.

### D. Motores de Videojuegos

El equivalente a la revolución industrial en la industria del videojuego, fue la creación de los motores de videojuegos (en inglés *Game Engine*). Un motor de videojuego es el elemento mas importante y la pieza fundamental del complejo proceso de desarrollo que implica la creación de un videojuego [3].

El motor de juego, o videojuego, hace referencia a una serie de rutinas que permiten la representación de todos los elementos de un juego. Es el resultado de la conjunción de diversos elementos, tales como: un motor gráfico, un motor de sonido, un gestor de IA o un motor físico. Cada pieza de este motor es fruto de un desarrollo específico, y debe ser la herramienta clave del programador para ese campo concreto. Por ejemplo, un motor gráfico va a incluir funciones, drivers y código que permitan relacionar lo programado con su representación gráfica, sin necesidad de entrar en detalles de hardware o sistemas, esto mismo es aplicable al motor de sonido o al físico.

Por ello, un motor de juego es el conjunto de herramientas con las que el programador o desarrollador va a trabajar para dar forma a su producto: el videojuego. Para que se pueda fabricar un videojuego hay que contar por tanto con un motor de juego y es en ellos donde las compañías desarrolladoras hacen sus grandes inversiones. Un buen motor de juego garantiza que durante un largo periodo de tiempo la empresa propietaria vaya a ser capaz de fabricar juegos con características y prestaciones actuales.

	Lumberyard	CryEngine	Unity 3D	Unreal Engine 4
Plataformas	PC, Xbox One, Playstation 4 (iOS, Android, Mac y Linux en progreso)	PC, Linux, Xbox One, Playstation 4	Todas las posibles	Windows PC, Playstation 4, Xbox One, Mac OS X, iOS, Android, Linux, SteamOS, HTML5
Soporte de Realidad Virtual	En progreso	En progreso	Excelente	Excelente
Importación de Assets	Necesidad de transformaciones	Compatibilidad muy reducida	Fácil, Gran cantidad de archivos soportados	Fácil
Lenguaje de Programación	C++, LUA	C#, C++	JavaScript, C#, Boo	C++
Ejemplos / Tutoriales	Casi inexistentes	Escasos	Amplios y variados	Amplios y variados
Facilidad de uso	No	No	Si	Depende del proyecto
Documentación	Poca, solo video tutoriales	Generalmente desfasada	Con todo lujo de detalles	Con todo lujo de detalles
Store	No existe	Muy reducido	Enorme	Buena
Programación Visual	No	Si, con Schematyc	Necesita de Addons	Si, con Blueprints
Networking	Basado en Servicios de Amazon	Requiere parámetros específicos	Unet	Fácil de implementar

Fig. 2. Tabla comparativa elaborada con el objetivo de comparar los distintos motores de videojuegos. Fuentes: [4] [5] [6] [7]

## III. PLANIFICACIÓN Y PRESUPUESTO

### A. Planificación

Principalmente se cuenta con cinco bloques de tareas diferenciados, englobados en el proyecto principal (numerado como 1):

2. *Revisión Bibliográfica*: Consta de 4 tareas: 2.1, 2.2, 2.3 y 2.4. Estas tareas consisten en la búsqueda de información y referencias sobre diversos campos como las herramientas a utilizar o las características y conceptos relativos a los videojuegos.

3. *Diseño e implementación (Videojuego)*: Engloba 9 tareas de la 3.1 a la 3.9. El objetivo de estos paquetes de trabajo es dar forma al videojuego del proyecto. Para ello hay tareas diversas, desde la iniciación en Unity hasta la creación de la interfaz gráfica del juego, pasando por el diseño de unidades o de controles del usuario.

4. *Diseño e implementación (Red)*: Esta tarea principal contiene diversas subtareas (4.1 a 4.11). En ellas se trabaja el apartado de red del proyecto, como el diseño de protocolos o la configuración del servidor que actúa como Nube.

5. *Pruebas*: Incluye 2 tareas: la 5.1 destinada a analizar el tráfico de red generada por el videojuego y la 5.2, cuyo objetivo es comparar el rendimiento de la aplicación creada con respecto a otros juegos similares.

B. *Presupuesto*:

#### IV. DISEÑO E IMPLEMENTACIÓN DEL VIDEOJUEGO

Material	Valor	Cantidad	Total
Ordenador Portátil Lenovo Ideapad 320	469'00 €	1	469'00 €
Smartphone Xiaomi Mi A1	181'50 €	1	181'50 €
Router inalámbrico - TP-Link N300 TL-WR841N	18'55 €	1	18'55 €
<b>Presupuesto Total Materiales:</b>			<b>669'05 €</b>

Fig. 3. Partida presupuestaria destinada a materiales.

Servicio	Valor	Cantidad	Total
Alquiler de Servidor VPS (Aruba Cloud)	1'00 € / mes	6	6'00 €
Tarifa de datos móviles 1,5Gbps (Simyo)	4'99 € / mes	6	29'94 €
Conexión a Internet de 100 Mbps (Movistar)	14'90 € / mes	6	89'40 €
<b>Presupuesto Total Servicios:</b>			<b>125'34 €</b>

Fig. 4. Partida presupuestaria destinada a la contratación de servicios.

Profesión	Valor	Cantidad	Total
Ingeniero Técnico de Telecomunicaciones	20'92 € / hora	825	17.259 €
Jefe de Proyecto - Contratado Doctor	29'07 € / hora	54	1569,78 €
<b>Presupuesto Total Personal</b>			<b>18.828'78 €</b>

Fig. 5. Presupuesto de personal necesario.

Partida	Valor
Presupuesto Materiales	669'05 €
Presupuesto Servicios	125'34 €
Presupuesto Personal	18.828'78 €
<b>Presupuesto Total Proyecto:</b>	<b>19.623'17 €</b>

Fig. 6. Presupuesto completo del proyecto.

A. *Aprendizaje previo*:

Como paso previo a la programación del videojuego resulta conveniente realizar un repaso a los conocimientos aprendidos durante el grado, así como a la documentación oficial del motor de juego Unity. El material de apoyo y los tutoriales disponibles en la red son del mismo modo una

valiosa fuente de información y permiten, antes de comenzar con el trabajo, comprender ciertos conceptos asociados a los videojuegos, el lenguaje a utilizar y el motor Unity.

Toda la documentación de Unity es accesible como recurso online a través de su página web [8]. Aunque la mayoría de la misma se encuentra disponible en castellano, es recomendable consultarla en inglés, ya que sus ejemplos y explicaciones resultan en ciertas ocasiones más claras y extensas.

Hay una amplia variedad de conceptos relativos a funciones y capacidades del motor *Unity*. Dichos conceptos fundamentales, como los Assets, los GameObject o los Component deben ser comprendidos, estudiados y asimilados con profundidad de cara al desarrollo a realizar.

B. *Diseño: Requisitos*:

Una vez se han adquirido una serie de conocimientos básicos sobre el motor de videojuego a utilizar, se está en disposición de comenzar con la labor de creación del mismo. No obstante, antes de lanzarse a programar conviene plantear cuales son las características deseadas para el videojuego. Con estas características en mente es posible comenzar a trabajar de forma ordenada, conociendo los objetivos finales a desarrollar.

Elementos a implementar:

- Pantalla/menu inicial de juego: con opciones de iniciar partida, salir o ajustar elementos gráficos.
- Interfaz de juego amigable y con un feedback de cara al usuario adecuado sin ser saturador.
- Capacidad de dos jugadores por partida.
- Sistema de juego basado en turnos con el referente del estilo de Panzer General.
- Existencia de un modo multijugador en el mismo dispositivo, modalidad hotseat.
- Existencia de un modo multijugador en Red de Área Local (LAN).
- Existencia de un modo multijugador vía Internet.
- Desarrollo enfocado a multiplataforma: PC (Windows) y Android.
- Capacidad de interacción entre versiones de distinta plataforma. Crossplay/juego cruzado entre Windows y Android.
- Guardado de partida y posibilidad de restaurar la misma con posterioridad.
- Capacidad de almacenamiento y carga de partidas situadas en nube remota.

Con estas características como objetivos, el diseño lógico del videojuego seguiría la estructura planteada en la Fig. 7. En esta figura se puede ver, a grandes rasgos, los elementos que componen el videojuego, así como la relación entre ellos y las clases que les dan forma.

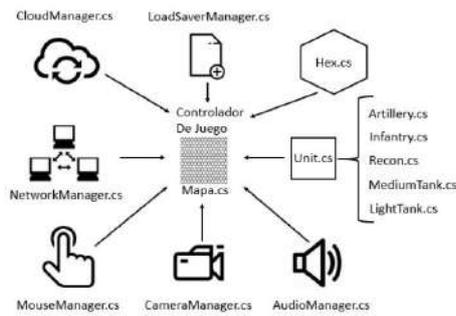


Fig. 7. Relación entre las principales clases del videojuego.

Se aprecia como el tablero de juego se encuentra formado por casillas, cada una con entidad y clase propia. Estas casillas y las unidades de juego, se encuentran ligadas al mapa o controlador de juego, el cual aglutina y sirve de nexo para el resto de clases y componentes del juego: como la interfaz gráfica, el controlador de sonido o el controlador de red. En el siguiente punto, es posible ver en mas detalle la utilidad del controlador de red, la clase Network Manager.

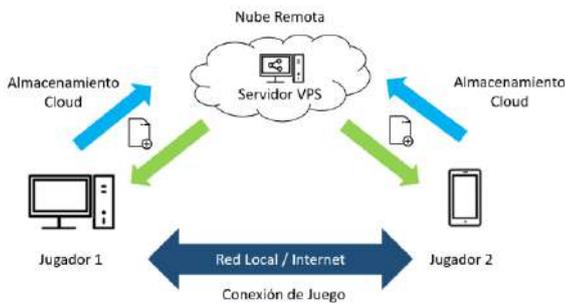


Fig. 8. Diseño de la parte en Red del proyecto.

En la Fig. 8, podemos ver la conectividad del videojuego. El juego admite partidas de dos jugadores, tanto en red local como mediante el uso de Internet. Así mismo, en la misma Fig. 8, se encuentran representadas las acciones de guardado remoto implementadas en un Servidor VPS Externo. Todos estos procesos hacen uso de clases creadas para tal fin como NetworkManager o CloudManager.

C. Áreas de trabajo:

El desarrollo del videojuego, se dividió en diversas áreas de trabajo, en su elaboración no se han utilizado añadidos externos de ningún tipo (excepto la apariencia de las fuentes de texto), el objetivo era realizar un proyecto íntegro, elaborando para el mismo todo lo necesario sin recurrir a ningún material externo. Las áreas de trabajo tratadas fueron las siguientes:

- Tablero de juego: Destacar la implementación de un algoritmo de pathfinding.
- Objetivo del juego: Ciudades. Casillas del juego decisoras del resultado del mismo.
- Concepto de jugador, cambio de turno. El jugador como ente indefinido, sistema basado en turnos.
- Unidades: diseño y acciones. Objetos interactuables, balanceo y esquemas claves de la acción in-game.
- Controles de Usuario.

- Apartado Gráfico: diseño de unidades y tablero.
- Interfaz y Menú de Juego: menú de configuración, opciones de resolución, apariencia e interfaz.
- Sonido.
- Guardado y carga Local: Aspecto a destacar, archivos de guardado propios con la información justa y necesaria para restaurar una partida.
- Juego en Red y Guardado en Nube: Tratado en el apartado 5.
- Distribución y Publicidad: Entendiendo el proyecto como un producto “completo” se realizaron estudios del mercado, así como el inicio de mecanismos publicitarios como el sitio web <https://www.cross-general.tk/>.

V. DISEÑO E IMPLEMENTACIÓN DEL PROTOCOLO

A. Protocolo General:

El videojuego, en su vertiente multijugador, cuenta con dos jugadores originarios de dos posibles plataformas, Android o PC. El uso de cualquiera de estas dos plataformas es indistinto de cara al diseño de los protocolos de comunicación, ya que el nivel de abstracción necesario no se ve afectado por estas características.

El primer paso a realizar para poder jugar una partida multijugador, es establecer una conexión entre un jugador A y un jugador B, para tal fin, se cuenta con un protocolo de conexión (y desconexión) que se utilizará tanto al principio como al final de cada partida (suponiendo una situación ideal sin caídas de conexiones). Dependiendo del escenario, este protocolo tendrá diferentes variantes: una enfocada a una Red de Área Local (LAN) y otra enfocada a una red pública como Internet.

Una vez se haya establecido la conexión, es el turno del protocolo de *Juego*. Este protocolo cuenta con las funciones necesarias para ir notificando el desarrollo de la partida a ambos jugadores, encargándose de transmitir la información asociada al movimiento de unidades, el combate o la captura de ciudades.

En cualquier momento de la partida puede hacerse uso del protocolo de *guardado*. En este caso, se establece una conexión entre uno de los dos jugadores (A o B) con un servidor VPS (Virtual Private Server) contratado para servir como *Nube* donde efectuar el almacenamiento remoto.

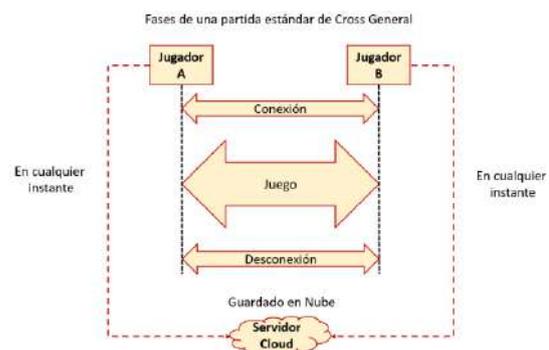


Fig. 8. Diseño de la parte en Red del proyecto.

Cuando la partida se da por finalizada, las funciones del protocolo de conexión son las encargadas de cerrar la comunicación. El esquema que seguiría una partida normal, puede verse en la Fig.9.

#### B. Implementación – Network Manager:

El motor de juego *Unity* cuenta con añadidos y utilidades propias como el *Network Manager* para gestionar las conexiones de red, sin embargo, aunque resulta enormemente útil para el usuario estándar, su configuración avanzada es sumamente limitada y no permite alterar u ofrecer nuevas formas de comunicación o conexión. Por este motivo, se ha optado en este proyecto por desarrollar un gestor de red o *Network Manager* propio.

Mediante al apoyo de librerías de *C#* enfocadas a las conexiones de red, se han creado funciones que permiten dar forma al protocolo ideado para el videojuego. Estas funciones implementan funcionalidades como la conexión mediante IP y puerto, la creación de *sockets* para comunicación vía UDP o funciones RPC destinadas a actualizar de forma simultánea ciertos datos en todos los clientes.

#### C. Protocolo de Conexión:

El protocolo de conexión se encarga de todas las funciones destinadas a permitir la conexión y desconexión entre los dos jugadores presentes en un escenario multijugador.

Es posible, de modo general, encontrarse ante dos situaciones de conexión: una en la que ambos jugadores se encuentran dentro de la misma red privada y otra en la que estos se encuentran en redes diferentes, interconectadas estas mediante el uso de una supared o red pública como Internet.

Los escenarios y métodos de conexión posibles se resumen a continuación:

- *Red privada o Red de Área Local:* Arquitectura típica disponible en un domicilio particular o en redes privadas mas amplias como las de un evento de videojuegos.

- *Conexión Directa:* Los jugadores realizan la conexión entre ambos mediante el uso de una conexión directa mediante IP privada de uno de ellos.
- *Método de Conexión Automática:* Un sistema basado en el envío de mensajes de tipo *Broadcast* (mensajes cuyo objetivo es llegar a todos los miembros de una red) permite establecer la conexión de manera automática entre ambos jugadores.

- *Interconexión de redes mediante Internet:*

- *Conexión Directa:* Los jugadores realizan la conexión entre ambos mediante el uso de una conexión directa mediante la IP pública de la red a la que está asociada uno de ellos.
- *Servicio de Sincronización:* Un servidor público accesible vía Internet, permite almacenar a los jugadores su nombre de usuario e IP para su posterior obtención y conexión por parte de un jugador rival.

En cualquier caso, conviene destacar que se sigue un esquema de *clientehost*. No hay un servidor de juego como

tal, sino que uno de los jugadores actúa como *host* mientras que su rival se conecta a él como *cliente*.

#### D. Protocolo de Juego:

La comunicación entre jugadores se realiza en determinadas situaciones en las que hay información clave que notificar al rival. Notificar cada acción de un jugador a su contrario puede parecer lo correcto, pero en una situación de red, cada comunicación tiene un coste, sobre todo si contamos con que el videojuego puede estar desarrollándose en algún terminal *Android* sujeto a un contrato de datos móviles.

Por este motivo, solo se realiza una comunicación entre jugadores en los siguientes casos: movimiento de una unidad, combate entre dos unidades, captura de una ciudad y cambio de turno. Las funciones destinadas a realizar estas comunicaciones se encuentran definidas en la clase *Network Manager* y reciben los nombres de *CheckUpdateMove*, *CheckUpdateCombat*, *UpdateCapture* y *UpdateTurn* respectivamente. Estas funciones son y están consideradas funciones RPC (*Remote Procedure Call*), es decir, son funciones o procedimientos que se activan o llaman de manera remota.

Como en el escenario de red planteado, no existe un servidor en sí, el valor de red, o la autoridad de cliente y host es la misma y solo se diferencian en que uno de ellos es el que se presta a albergar la partida o la conexión. Estas funciones RPC, pueden ser ejecutadas allí donde estén definidas. En el caso de este proyecto, como las versiones del programa son exactamente las mismas para los dos jugadores, cuando un jugador hace uso de una función RPC, puede *invocar* o llamar a esa misma función en el lado del jugador rival.

#### E. Almacenamiento en la Nube:

En primer lugar, para poder implementar un servicio de *Cloud Storage*, es necesario contar con un dispositivo o equipo con conexión a Internet que haga las veces de servidor. En un primer momento se pensó en utilizar un PC obsoleto o una *Raspberry Pi*, sin embargo, se disponen de múltiples soluciones en línea que pueden ofrecer un mejor resultado.

Dentro de estas soluciones comerciales, se ha optado por elegir un servidor VPS (Virtual Private Server) de la empresa *Aruba Cloud*. La dirección pública del equipo es la 188.213.166.227, también es posible acceder al mismo mediante un dominio adquirido y configurado para tal fin: [cross-general.tk](http://cross-general.tk)

Para lidiar con el servicio de almacenamiento en nube se cuenta con una clase creada específicamente para tal fin: *CloudManager*. Esta clase contiene dos únicas funciones: *CloudUpload* y *CloudDownload*, la primera destinada a la subida de archivos y la segunda destinada a la bajada de los mismos. La función *CloudUpload* realiza una conexión web entre el cliente y el servidor VPS de almacenamiento. Esta conexión apunta a un archivo PHP concreto del servidor virtual, creado para implementar el almacenamiento en la nube.

El archivo PHP contiene la reacción que debe ofrecer el

servidor ante una petición HTTP de tipo POST. El código presente en el archivo PHP hace que el servidor solo acepte aquellos archivos con el formato de guardado del videojuego, es decir, aquellos archivos con terminación *.sav*. Además, el archivo PHP, incluye el directorio final donde la partida a subir será finalmente almacenada (*/var/www/html/cloudsave/savegames/*).

En el lado del cliente, además de crear la conexión web, es necesario captar todos los posibles errores que pueda generar la conexión web o la petición HTTP. También antes de enviar el archivo, es necesario generarlo tal y como se hace para un guardado local. Por último, mencionar que para su transmisión, el archivo es transformado en un vector de bytes.

## VI. ANÁLISIS Y RESULTADOS

Tras el desarrollo del videojuego, y durante el mismo, se han realizado pruebas con distintos dispositivos y escenarios, con el fin de comprobar el correcto funcionamiento del programa elaborado. Además de analizar el tráfico de red generado por el videojuego, se realiza una comparación de su rendimiento con el de un título similar disponible para *Android*.

Longitud (bytes)	Tipo
57	ACK
60	ACK
74	Connected Ping / Pong
66	Connection Request
69	Connection Request Accepted
1506	New Incoming Connection

Fig. 9. Tipo de mensajes del protocolo RakNet identificados según su longitud en bytes.

Destacar el análisis y la identificación de tráfico del protocolo RakNet que trabaja sobre UDP y sobre el cual, los analizadores de paquetes no reflejaban información alguna. Gracias a numerosas pruebas de cotejo y relación de mensaje, se han logrado identificar patrones relacionados con diferentes procesos de la comunicación. Así mismo, los mensajes intercambiados en dichos patrones han sido igualmente identificados a tenor de su intercambio, función y número de bytes que lo componían.

## VII. CONCLUSIONES

Tras este desarrollo, se ha evaluado, gracias a los resultados de los análisis de tráfico, que el sistema implementado cumple con los requisitos de diseño que se establecieron. Así mismo, Comparando su rendimiento con títulos similares, se aprecia que tanto su consumo de batería, como de datos, entra dentro de los marcos habituales para un juego de sus características.

Por tanto, mediante este desarrollo, se ha podido comprobar, que es posible realizar juegos inter-plataforma, sin que esto suponga un desafío mucho mayor al de desarrollar un videojuego tradicional. Se dispone de grandes herramientas para ello y la única precaución a tener en cuenta, es ser consciente desde un primer momento, de las plataformas de juego objetivo.

Adicionalmente, también se ha podido comprobar que implementar un servicio *Cloud* es posible con muy pocos recursos, conviene recordar, que el servidor utilizado supone un coste de solo 1 euro al mes. Así mismo, se ha comprobado que un pequeño desarrollador, o equipo de trabajo, puede lanzar al mercado su producto y hacer publicidad del mismo con un coste mínimo. Todo ello gracias a las plataformas de distribución existentes, como *Steam* o *Google Play* y al auge de las redes sociales.

Como resumen final, se exponen a continuación las tareas o logros alcanzados en la realización del proyecto:

- Estudio del género de videojuego elegido, del mercado actual y de las soluciones existentes que hacen uso de *Cloudsaving* o de *Crossplay*.
- Estudio y comparación de las herramientas disponibles para el desarrollo de videojuegos. Aprendizaje del motor de videojuegos *Unity* y el lenguaje de programación *C#*.
- Diseño, desarrollo e implementación de un videojuego TBS multiplataforma.
- Diseño de un protocolo que permite el *Crossplay* entre jugadores de distintas plataformas.
- Diseño de un protocolo y arquitectura que permite el *Cloudsaving*.
- Configuración de un Servidor Web para que ofrezca servicios de almacenamiento remoto.
- Realización de pruebas de validación en diferentes escenarios tanto de la vertiente multijugador como del funcionamiento del *Cloudsaving*.
- Estudio y análisis del tráfico de datos y del consumo del videojuego. Simulación y estudio de labores de distribución y publicidad

## AGRADECIMIENTOS

A mi familia, pareja y amigos, ha sido un largo camino que sin vuestra ayuda no hubiese podido recorrer.

A mi tutor Juanjo por darme la oportunidad de dar forma a algo que siempre he deseado hacer.

Gracias a todos.

## REFERENCIAS

- [1] Statista. Number of players of Fortnite worldwide from August 2017 to June 2018 (in millions). <https://www.statista.com/statistics/746230/fortnite-players/>, Junio 2018.
- [2] Wenyng Zeng, Yuelong Zhao, Kairi Ou, and Wei Song. Research on cloud storage architecture and key technologies. In *Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human*, pages 1044-1048. ACM, 2009.
- [3] David Vallejo Fernandez, Cleto Martin Angelina, et al. *Desarrollo de Videojuegos: Arquitectura del Motor de Videojuegos*, volume 1. Cursos en Español, 2011.
- [4] Sergey Korolev. *How to Make Lumberyard Better?* <https://80.lv/articles/opinion-how-to-make-lumberyard-better/>, Febrero 2017.
- [5] Crytek. *Sitio Web Oficial del Motor Cryengine*. <https://www.cryengine.com/>, 2018.
- [6] Amazon. *Sitio Web Oficial del Motor Lumberyard*. <https://aws.amazon.com/es/lumberyard/>, 2018.
- [7] Epic Games. *Sitio Web Oficial del Motor Unreal Engine*. <https://www.unrealengine.com/en-US/blog>, 2018.
- [8] Unity Enterprise. *Sitio web oficial del Motor de Videojuego Unity*. <https://unity3d.com/es>, 2001-2018.

# ***Teoría de la Señal y Comunicaciones***



# Diseño de antena en array basada en tecnología SIW para milimétricas

Autor: Bilal Hammu Mohamed, e-mail: gemelo98@correo.ugr.es

Tutor: Pablo Padilla de la Torre, pablopadilla@ugr.es

Cotutor: Francisco Javier García Ruiz, franruiz@ugr.es

Titulación: Ingeniería de Tecnologías de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—El ingente flujo de datos producido por una sociedad totalmente digitalizada exige mayores velocidades de transmisión. Esto, junto al mayor número de dispositivos conectados al mismo tiempo, requiere elevados anchos de banda. Por este motivo, el diseño de dispositivos de radiofrecuencia tiende a utilizar frecuencias cada más elevadas, permitiendo anchos de banda mayores y menores ratios de Señal-Interferencia

En los sistemas de comunicación actuales uno de los factores más críticos y con mayores líneas de investigación es la optimización del funcionamiento de las antenas, tratando de dotarlas de mayor versatilidad en sus parámetros de radiación, así como de reconfigurabilidad sin perjudicar las especificaciones de la antena.

El presente proyecto consiste en el diseño, fabricación y caracterización de una agrupación de antenas usando la tecnología Substrate Integrated Waveguide (SIW) en la banda de 35 a 40 GHz. Además, se plantea una solución para modificar electrónicamente la dirección de apuntamiento de la antena.

**Palabras clave**—Agrupación, Ancho de banda, Antena, Desfasador, Directividad, Fase, Formación de haz, Ganancia, Guías de onda, IoT, Radar, Ranura, SIW, tracking.

## I. INTRODUCCIÓN

EL gran desarrollo de las tecnologías de la información y las comunicaciones (TIC) hace necesario un aumento en la velocidad de transmisión de datos lo que requiere de una evolución de las redes de telefonía móvil permitiendo la utilización de anchos de banda cada vez mayores. Para lograr dicho aumento en las velocidades de transmisión se trata de lograr antenas que funcionen a frecuencia cada vez más elevadas [1], [2]. A pesar de las ventajas que puede otorgar trabajar en rangos de frecuencia mayores también puede suponer ciertos inconvenientes, como un aumento en las pérdidas de energía. Este efecto se trata de paliar con el uso de antenas de mayor ganancia [3], [4].

Una de las posibles formas de lograr este aumento de la ganancia es mediante el uso de una agrupación de antenas [5], que consiste en colocar varias antenas iguales separadas una cierta distancia con el fin de modificar las características de radiación del conjunto de antenas.

Dada la enorme cantidad de dispositivos funcionando al mismo tiempo y las limitaciones en cuanto a ancho de banda es conveniente que los sistemas emisores busquen la ruta más eficiente para transmitir información e ir modificando dicha ruta según las situaciones que se den en el medio. A esto

se le conoce como beamforming [6], la cuál es una técnica utilizada para modificar la dirección de apuntamiento de las antenas con el fin de obtener una transmisión más eficiente. Para la realización de dispositivos de alta frecuencia una de las alternativas es la utilización de la tecnología SIW (Substrate Integrated Waveguide) [2] la cual ofrece a altas frecuencias mejores prestaciones que otras tecnologías como son la microstrip o las guías de ondas coplanares [7]. Dicha tecnología busca asemejarse a una guías de ondas utilizando un sustrato con una capa metálica superior e inferior que actuarán modelarán las paredes superiores e inferiores. Para modelar las paredes laterales se hará uso de cilindros metálicos colocados con una cierta separación.

Uno de los retos de la bibliografía actual para todar a los dispositivos de reconfigurabilidad es conseguir la realización de desfasadores controlables en guías de ondas a alta frecuencia sin sufrir un empeoramiento considerable en las prestaciones. Esta variación controlada de fase en una agrupación podría dar lugar a un beamformer, es decir, poder controlar la dirección de apuntamiento de la antena.

En el presente proyecto se planteará el diseño, fabricación y caracterización experimental de una agrupación de antenas con una alta ganancia en tecnología SIW cuyo funcionamiento estará centrado en la banda de 35 GHz y se buscará un ancho de banda lo más elevado posible.

Otro de los retos a lograr será la realización de un desfasador activo. Basándose en la propuesta de Ding y Wu [8] se tratará de lograr dicho objetivo mediante la realización de una ranura en la red de alimentación sobre la cuál se le colocará un varactor y mediante un control de su capacidad se modificará la fase con la que recibe la onda cada una de las salidas de dicha red, de forma que al alimentar cada antena con una fase distinta se logre un control de la dirección de máximo apuntamiento de la antena.

## II. ESTRUCTURA GENERAL DE LA ANTENA

En el presente proyecto se plantea un dispositivo basado en la tecnología SIW. Para ello, se hizo uso de dicha tecnología utilizando el sustrato Rogers R04003C [9] que posee una altura de 1.524 mm y permitividad eléctrica de  $\epsilon_r = 3.55$ . Los cilindros metálicos que se utilizarán para simular las paredes laterales de la guía de onda contarán con un diámetro de 0.4 mm y una separación de 0.7 mm entre cada uno de ellos.

Para formar la antena, nuestra estructura cuenta con dos capas, la capa inferior forma la red de alimentación siendo esta un divisor con una entrada y cuatro salidas mientras que la capa superior se compone de cuatro cavidades resonantes que serán alimentadas por el divisor de forma similar a como se planteó en [10]. En la imagen 1 se plantea un esquema del diseño general de la antena.

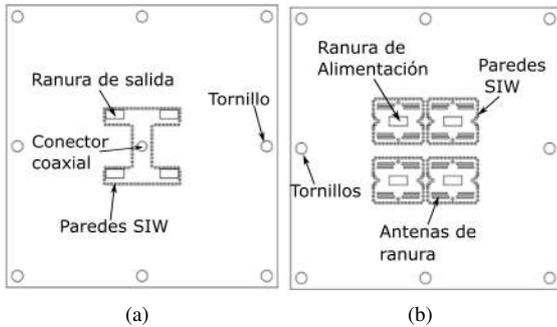


Imagen 1. Estructura de la antena (a) Divisor 1 a 4; (b) Cavidades.

En cuanto a la alimentación del divisor, de las múltiples formas que existen para alimentarlo [11], será alimentado a través de un conector coaxial en abierto. Dicho conector coaxial estará formado por un aislante de teflón de 1.67 mm de diámetro y un vivo de cobre de 0.5 mm de diámetro. En cuanto a las salidas del divisor, este cuenta con 4 ranuras de salida las cuales conectan con la capa superior.

En la segunda capa cada cavidad cuenta con una ranura de alimentación en su parte inferior con el fin de unirse con la ranura de salida del divisor. Además, cada una de las cavidades cuenta con cuatro parejas de ranuras de salida, las cuáles serán nuestros elementos radiantes. Al colocar cuatro cavidades cada una de ellas con cuatro parejas de ranura se forma una agrupación de antenas 4x4. Con el fin de unir dichas capas se hará uso de tornillos.

### III. DISEÑO DE LA ANTENA

El diseño se realizó centrado en la banda de frecuencias de 35 GHz. Para la realización de dicho proyecto se hizo uso del software CST Studio [12].

#### A. Cavidad resonante en tecnología SIW

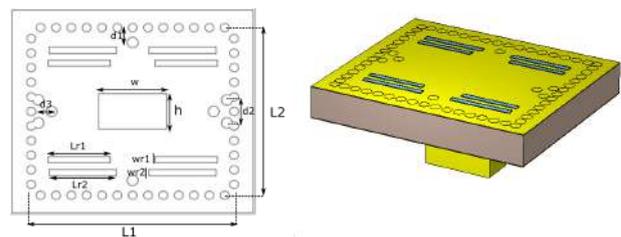
Dicha cavidad será una cavidad resonante rectangular la cual cuenta con una ranura en la parte inferior colocada en el centro de esta a través de la cual es alimentada. En la parte superior cuenta con cuatro parejas de ranuras de diferente longitud. Dado que una ranura se encuentra en resonancia cuando la longitud de esta es aproximadamente  $\frac{\lambda}{2}$  [13] al colocar dos ranuras, puesto que estas tendrán frecuencias de resonancia próximas, se puede lograr un ancho de banda superior a lo que se lograría con solo una ranura.

En la Imagen 2 se aprecia el diseño de la cavidad. En ella es apreciable como la cavidad no es completamente rectangular sino que cuenta con unos entrantes a los extremos. Ya que dichos entrantes son metálicos harán el mismo efecto que haría una pared y por tanto el campo electromagnético no es capaz de distribuirse por aquella zona lo que modifica la distribución de campos en el interior de la cavidad y puede

facilitar una adaptación de impedancias.

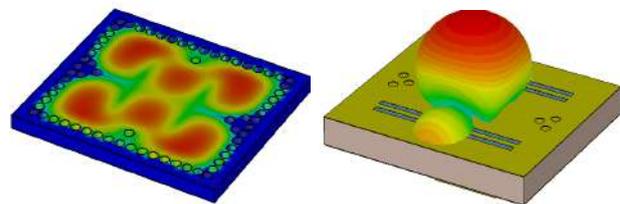
En la Imagen 3 se muestra el módulo del campo eléctrico en el interior de la cavidad así como el diagrama de radiación de esta. Observando el campo eléctrico se aprecia como efectivamente los cilindros modifican la distribución electromagnética en el interior de la cavidad. Por otro lado, en el diagrama de radiación es apreciable como la radiación de la cavidad tiene una tendencia direccional en la dirección perpendicular a esta.

El módulo del coeficiente de reflexión o parámetro  $|S_{11}|$  de la cavidad se muestra en la imagen 4. Se considera como aceptable todo valor de frecuencias para el que el módulo del coeficiente de reflexión esté por debajo de  $-10$  dB, por tanto en este caso nuestra cavidad tendrá un ancho de banda de aproximadamente 2 GHz y estará centrada en el valor de 35 GHz.



(a) Planos de la cavidad resonante. (b) Vista tridimensional de la cavidad resonante.

Imagen 2. Cavidad resonante (a) Planos de la cavidad; (b) Vista 3D de la cavidad.



(a) Módulo del campo eléctrico a 35 GHz (b) Diagrama de radiación de la cavidad a 35 GHz.

Imagen 3. (a) Simulación del módulo del campo eléctrico en el interior de la cavidad a 35 GHz; (b) Diagrama de radiación de la cavidad a 35 GHz.

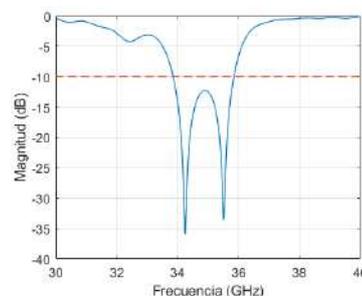
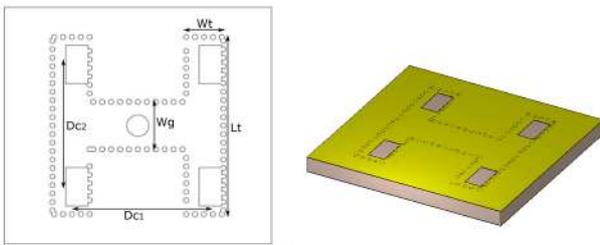


Imagen 4. Simulación  $|S_{11}|$  de la cavidad.

#### B. Alimentación de las cavidades: Divisor 1 a 4

La alimentación de la antena se realizó mediante un divisor 1 a 4 que distribuya la energía electromagnética por todas las



(a) Planos del divisor. (b) Vista tridimensional del divisor

Imagen 5. Divisor 1 a 4 (a) Planos del divisor; (b) Vista 3D del divisor.

antenas. Este es alimentado con un conector coaxial en la parte inferior y cuatro ranuras de salida. Dicha estructura debe de funcionar en el mismo rango de frecuencias que la cavidad, es decir, en torno a 35 GHz.

En la Imagen 5 se muestra el plano del divisor realizado así como una imagen tridimensional. El ancho de banda del divisor y su frecuencia de funcionamiento viene determinado por el parámetro  $|S_{11}|$ . Es conveniente que el rango en el que dicho parámetro sea óptimo sea superior al de la cavidad, así la red de alimentación no será la que limite el ancho de banda.

Debido a que esta estructura cuenta con varios puertos de salida es necesario estudiar los coeficientes S en transmisión, o los parámetros  $|S_{n1}|$ , los cuales representa la potencia que recibe el puerto n respecto a la potencia entrante. Al tener 4 puertos de salida, idealmente dicho parámetro tomará el valor de  $-6$  dB en el rango de funcionamiento. En la Imagen 6 se muestra el coeficiente de reflexión así como los parámetros S de transmisión. Observando los parámetros S se observa como el rango de frecuencias de funcionamiento es el mismo que el de la cavidad y además el ancho de banda que posee el divisor es superior, por tanto no presenta una limitación en el rango de frecuencias disponible.

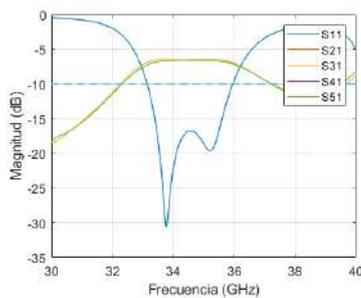


Imagen 6. Simulación del parámetro  $|S_{n1}|$  del divisor 1 a 4.

### C. Antena completa

Tras la realización de ambos componentes, el divisor y las cavidades, estos deben de ser unidos para formar la antena completa. En la Imagen 7 se muestra el prototipo de la antena completa, además de una imagen de la cámara anecoica dónde se realizaron las medidas

El correcto funcionamiento de la estructura vendrá determinado con el coeficiente  $|S_{11}|$  y por el el diagrama de radiación. Este apartado consta de un análisis de las características de la antena completa así como una

comparativa entre los resultados experimentales y los obtenidos en simulación, además de un estudio de los motivos que generan las posibles discrepancias.

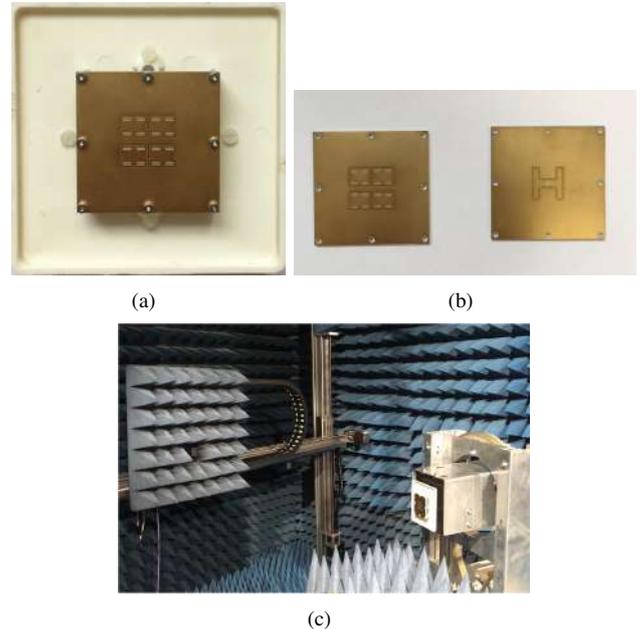


Imagen 7. Prototipo de la antena completa (a) Antena completa; (b) Divisor y cavidades fabricados; (c) Medidas en cámara anecoica.

1) *Estudio del coeficiente de reflexión:* En la Imagen 8 se muestra una comparativa entre el coeficiente  $|S_{11}|$  del prototipo de la antena completa y la simulación de dicho parámetro. Cabe destacar las apreciables diferencias que presentan, variando el valor de frecuencia central de 35 GHz a 36 GHz. Además hay una reducción considerable del ancho de banda pasando de 2.3 GHz a 1.8 GHz.

La modificación de la frecuencia central puede deberse a multitud de factores, entre ellos las posibles variaciones en las características del sustrato. En la Imagen 9 se muestra las variaciones que se dan en el coeficiente de reflexión al realizar la simulación con distintos valores de permitividad eléctrica  $\epsilon_r$ . En estas se puede observar como al reducir dicho valor de 3.55 a 3.35 el rango de frecuencias de funcionamiento de la antena se asemeja al resultado obtenido experimentalmente, por tanto se confirma que las diferencias vienen dadas por variaciones de las posibles propiedades del sustrato.

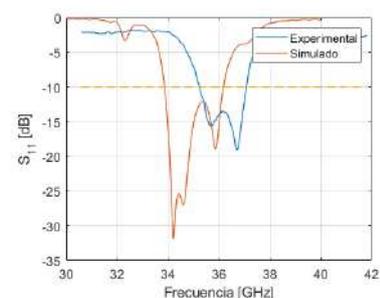


Imagen 8. Comparativa del  $|S_{11}|$  experimental y simulado

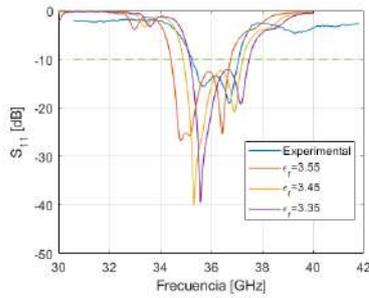


Imagen 9. Comparativa  $|S_{11}|$  simulado para diferentes valores de  $\epsilon_r$  y experimental.

2) *Estudio del diagrama de radiación:* En la Imagen 10 se muestra una comparativa entre la simulación del diagrama de radiación obtenido en las simulaciones y el resultado experimental del prototipo completo. Además se muestra una vista superior del diagrama obtenido experimentalmente 36.6 GHz. Se aprecia similitud entre ambos resultados, obteniéndose para ambos una ganancia de aproximadamente 17.5 dBi y una antena con una tendencia directiva.

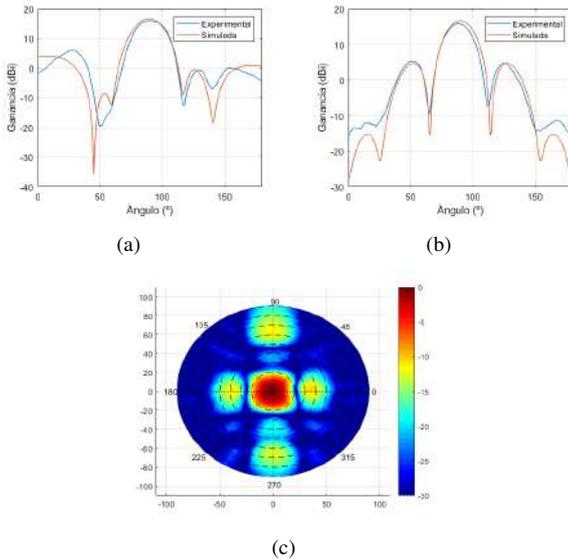


Imagen 10. Comparativa del diagrama de radiación obtenido mediante simulación y experimentalmente (a) Comparativa en plano E; (b) Comparativa en plano H; (c) Vista superior del diagrama de radiación.

Las dimensiones de la antena se muestran en la tabla I.

Tabla I  
MEDIDAS DE LOS DISTINTOS PARÁMETROS DE LA ANTENA

Parámetros	Medidas (mm)	Parámetros	Medidas (mm)
$L_1$	9.26	$d_1$	0.6
$L_2$	7.96	$d_2$	0.38
$w$	3.2	$d_3$	1.0
$h$	1.45	$Dc_1$	10.5
$Lr_1$	2.84	$Dc_2$	10.07
$Lr_2$	3.07	$W_g$	3.65
$wr_1$	0.284	$W_t$	2.9
$wr_2$	0.307	$L_t$	13.95

#### IV. DESFASADOR ACTIVO

La implementación de un desfaseador activo en una guía de onda que sea controlable y a la vez mantenga unas

prestaciones adecuadas en cuanto a pérdidas de potencia y ancho de banda es uno de los problemas aún sin resolver en la bibliografía actual. En el presente proyecto se trató de realizar una variación en la red de alimentación de la antena presentada previamente con el fin de que actúe como desfaseador.

Para la realización del desfaseador se modificó el divisor 1 a 4 presentado previamente colocándole en su parte inferior una serie de ranuras con el objetivo de introducir un varactor en éstas y así modificar la capacidad equivalente de dicho varactor generando una diferencia de fases entre unos puertos de salida y otros. En la Imagen 11 se muestra una vista inferior del divisor con las ranuras mencionadas donde se aprecia como se han colocado una pareja de ranuras a cada extremo del varactor, una de ellas en lo nombrado como *zona 1* y otro en la *zona 2*. La numeración mostrada se refiere al número con el que se identifica cada uno de los puertos encontrados en dicha zona.

Se puede ver como en cada extremo se realizaron dos ranuras, esto se hace con el objetivo de colocar dos varactores en lugar de uno y así generar una mayor variación respecto a la diferencia de fase. Cabe destacar la relevancia que tienen la posición de las ranuras así como sus dimensiones ya que dichas variables pueden modificar el desfase que se da así como la frecuencia a las que se da este desfase. En este caso la longitud de la ranura de mayor tamaño es de 5.9 mm y cuenta con un grosor de 0.15 mm. La ranura de menor tamaño tiene una longitud de 5.5 mm y un grosor de 0.1 mm. En cuanto a la posición de las distintas ranuras,  $D = 4.9$  mm,  $D_{var} = 1$  mm.

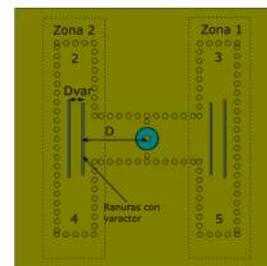


Imagen 11. Parte inferior del divisor con las ranuras para los varactores.

Para el estudio del desfase generado, dado que el puerto 2 presenta simetría con el puerto 4 y a su vez el puerto 3 presenta simetría con el 5 sería suficiente con estudiar la diferencia de fase generada entre el puerto 2 y el puerto 3. En la Imagen 12 se muestra  $\phi_2 - \phi_3$  donde  $\phi_2$  y  $\phi_3$  representan la fase respecto al puerto 1 del puerto 2 y 3 respectivamente. En esta se observa como el desfase máximo está en torno a  $-120^\circ$ . Aunque este desfase se da en un valor de frecuencias ligeramente superior al que se definió como rango óptimo en la simulación de la antena, esto no supondría ningún inconveniente ya que como se observó con la fabricación tiende a aumentar la frecuencia de funcionamiento haciendo que la frecuencia a la que se da el desfase y la frecuencia de funcionamiento coincidan.

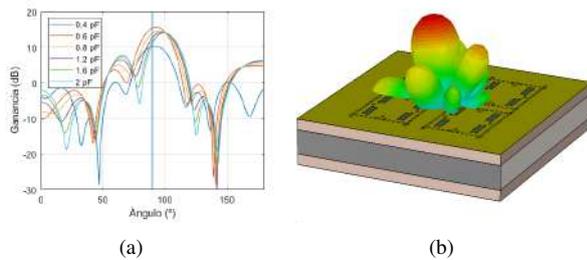


Imagen 13. (a) Simulación del diagrama radiación para distintos valores de capacidad; (b) Vista tridimensional del cambio en la dirección de apuntamiento.

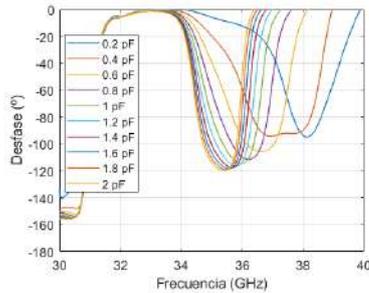


Imagen 12. Desfase generado entre el puerto 2 y el puerto 3 a distintas frecuencias para distintos valores de capacidad.

Lo esperable sería que la colocar las cavidades sobre el divisor se generase una diferencia de fase entre las cavidades lo que modificaría la dirección de apuntamiento. Sin embargo, el campo electromagnético no se distribuye de forma uniforme por el puerto de salida del divisor, lo que genera diferencias de alimentación entre las ranuras de una misma cavidad y el efecto en el apuntamiento no es el deseado. Para paliar esto es necesario la colocación de una guía de onda de transición entre las ranuras del divisor y la ranura de las cavidades de forma que haga simétrica la alimentación de las ranuras de una misma cavidad y se logre un mejor desfase. En la imagen 13 se muestra la simulación del diagrama de radiación en plano E del divisor desfasador con cuatro cavidad, una en cada puerto de salida. Dicho diagrama de radiación se muestra para distintos valores de capacidad para así mostrar como la dirección de máximo apuntamiento varía con la capacidad. Efectivamente se ha logrado la variación de apuntamiento deseada puesto que hay una modificación progresiva según aumenta la capacidad consiguiéndose una variación máxima de  $10.5^\circ$ .

Un inconveniente que se observa es una reducción de la directividad, mientras que en el apartado III-A esta se encontraba en torno a 18 dB, en este caso se encuentra en torno a 14 dB. Dicha reducción se debe al efecto generado por las ranuras del desfasador ya que generan que parte de la potencia se radie por estas. A pesar del correcto funcionamiento a la hora de generar el desfase, este diseño cuenta con una serie de inconvenientes ya que hay una fuerte variación del parámetro  $|S_{11}|$  con la capacidad haciendo que el funcionamiento de la antena no sea óptimo. Además, las ranuras en las que se encuentran los varactores generan considerables fugas de potencia, de forma que gran parte de la energía no es radiada como se debería.

## V. AMPLIACIÓN DE LA ANTENA

Como se explicó, al realizarse una agrupación de antenas es posible modificar el diagrama de radiación. Esto se puede utilizar con el fin de aumentar la directividad lo cual se logra con un mayor número de antenas.

Por ello, con el fin de lograr una mayor directividad se planteará una ampliación de la antena presentada en II.

Para alcanzar este propósito, se volverá a rediseñar la antena de forma que en la primera capa se realizará un divisor 1 a 4 similar al planteado en III-B pero de un mayor tamaño.

En la segunda capa se colocarán cuatro divisores similares al planteado anteriormente con la excepción que estos estarán alimentados por el divisor de la capa inferior a través de una ranura.

Finalmente, cada uno de los divisores de la segunda capa tendrán en sus salidas la tercera capa dónde se encontrarán cavidades resonantes idénticas a la planteadas en III-A. Con esto se logrará una agrupación de 16 cavidades y puesto que cada cavidad cuenta con cuatro parejas de ranuras, esto da lugar a una agrupación  $8 \times 8$ . En la Imagen 14 se muestran los distintos componentes además de la antena completa.

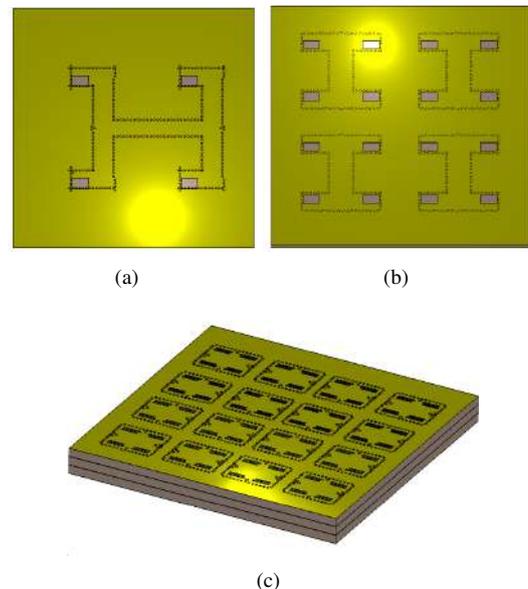


Imagen 14. Antena ampliada (a) Vista superior de la capa inferior; (b) Vista superior de la capa intermedia; (c) Vista en perspectiva de la antena completa.

El módulo del coeficiente de reflexión aquí logrado se muestra en la Imagen 15 dónde se ve como funciona en la frecuencia central de 35 GHz con un ancho de banda de aproximadamente 2 GHz. Con el fin de comprobar la mejora de la directividad derivada de aumentar el número de antenas en la Imagen 16 se muestra una comparativa del diagrama de radiación de una agrupación de una cavidad, cuatro cavidades y 16 cavidades. Además se muestra una vista tridimensional del diagrama de radiación de la antena completa.

Dado que con una cavidad se consiguió una directividad de en torno a 12 dBi, con cuatro cavidades esta era de aproximadamente 18 dBi y con 16 cavidades de 24 dBi. Por tanto aprecia un claro aumento de la directividad conforme aumenta el número de antenas.

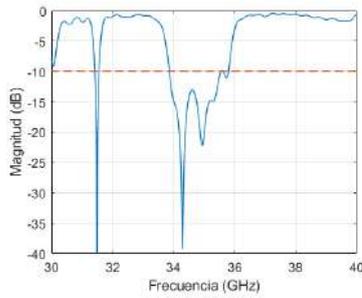


Imagen 15. Simulación del  $|S_{11}|$  de la antena ampliada.

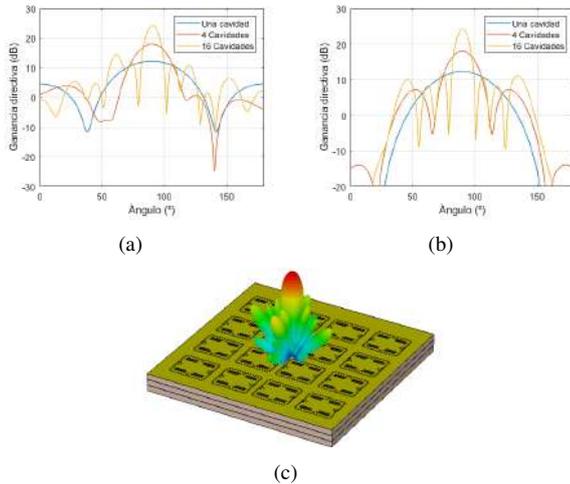


Imagen 16. Comparativa del diagrama de radiación obtenido para una cavidad, cuatro cavidades y dieciséis cavidades (a) Comparativa en plano E; (b) Comparativa en plano H; (c) Vista tridimensional de el diagrama de radiación de la antena completa.

## VI. CONCLUSIONES

Este documento se puede dividir en tres secciones:

- Primeramente se realizó una agrupación de antenas  $4 \times 4$  en tecnología SIW la cuál cuenta con un ancho de banda de 1.8 GHz en la banda de 36 GHz. A su vez se logró una ganancia en torno a 17.5 dB.
- Con el fin de variar la dirección de apuntamiento se modificó la red de alimentación para conformar un desfase controlable. Se logró un desfase máximo entre puertos de  $120^\circ$  y una diferencia en la dirección de máximo apuntamiento de  $10.5^\circ$ .
- Buscando una mayor ganancia se planteó realizar una mayor agrupación de un número mayor de antenas. Para ello se aumentó la red de alimentación y se realizó una agrupación de  $8 \times 8$  antenas. La ganancia de dicha antena es de  $21 \text{ dBi}$  y funciona a una frecuencia central de 35 GHz con un ancho de banda de 2 GHz.

## AGRADECIMIENTOS

Mi más sincero agradecimiento a mis tutores Pablo Padilla de la Torre y Francisco Javier García Ruiz por toda la ayuda ejercida como tutores, además de permitirme realizar las mediciones en el laboratorio de 5G del CITIC.

## REFERENCES

- [1] IEEE. URL: <https://spectrum.ieee.org/video/telecom/wireless/everything-you-need-to-know-about-5g>.
- [2] A.R Hernandez. *Capítulo 4 Substrate Integrated Waveguide (SIW)*. URL: [http://catarina.udlap.mx/u\\_dl\\_a/tales/documentos/lem/hernandez\\_a\\_r/capitulo4.pdf](http://catarina.udlap.mx/u_dl_a/tales/documentos/lem/hernandez_a_r/capitulo4.pdf).
- [3] D. Lockie and D. Peck. “High-data-rate millimeter-wave radios”. In: *IEEE Microwave Magazine* 10.5 (2009), pp. 75–83.
- [4] B. He, Y. Jiao, and H. He. “Design of high-gain lens antenna for 5G application”. In: *2017 Sixth Asia-Pacific Conference on Antennas and Propagation (APCAP)*. 2017, pp. 1–3.
- [5] Constantine A Balanis. *Antenna theory: analysis and design*. John Wiley & sons, 2016.
- [6] Q. Xu et al. “Waveforming: An Overview With Beamforming”. In: *IEEE Communications Surveys Tutorials* 20.1 (2018), pp. 132–149.
- [7] M. Bozzi, L. Perregini, and Ke Wu. “Modeling of losses in substrate integrated waveguide by Boundary Integral-Resonant Mode Expansion method”. In: *2008 IEEE MTT-S International Microwave Symposium Digest*. June 2008, pp. 515–518. DOI: 10.1109/MWSYM.2008.4633216.
- [8] Y. Ding and K. Wu. “Varactor-tuned substrate integrated waveguide phase shifter”. In: *2011 IEEE MTT-S International Microwave Symposium*. 2011, pp. 1–4.
- [9] Rogers. *RO4000 Series*. URL: <https://rogerscorp.com/-/media/project/rogerscorp/documents/advanced-connectivity-solutions/english/data-sheets/ro4000-laminates-ro4003c-and-ro4350b---data-sheet.pdf>.
- [10] J. Xu et al. “Bandwidth Enhancement for a 60 GHz Substrate Integrated Waveguide Fed Cavity Array Antenna on LTCC”. In: *IEEE Transactions on Antennas and Propagation* 59.3 (2011), pp. 826–832.
- [11] P.A. Rizzi. *Microwave Engineering: Passive Circuits*. Passive Circuits. Prentice Hall, 1988. ISBN: 9780135867020. URL: <https://books.google.es/books?id=h5hrQgAACAAJ>.
- [12] CST Studio. *CST STUDIO*. URL: <https://www.3ds.com/es/productos-y-servicios/simulia/productos/cst-studio-suite/>.
- [13] Miguel Ferrando. *Ranuras y Antenas microstrip*. URL: [http://www.upv.es/antenas/Documentos\\_PDF/Notas\\_clase/Antenas\\_microstrip.pdf](http://www.upv.es/antenas/Documentos_PDF/Notas_clase/Antenas_microstrip.pdf).



**Bilal Hammu Mohamed** Nació y creció en España, concretamente en la ciudad de Melilla. Realizó el grado en Ingeniería de Tecnologías de Telecomunicación en la Universidad de Granada. Actualmente se encuentra realizando el Máster en Ingeniería de Tecnologías de Telecomunicación en la Universidad de Granada.

# Diseño de antenas de bocina SIW para sistemas de medidas de antenas

Autor: Cleofás Segura Gómez, e-mail: cleofas@correo.ugr.es

Tutores: Pablo Padilla de la Torre, e-mail: padilla@ugr.es

Juan Francisco Valenzuela Valdés, e-mail: juanvalenzuela@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—Este documento presenta el diseño de una antena de bocina *Substrate-Integrated Waveguide* (SIW) basada en una apertura del plano H. La antena está compuesta por una alimentación coaxial y una apertura. En esta apertura se introducen vías metálicas que actúan como paredes parcialmente separadas para formar 4 sub-aperturas adyacentes al final de la apertura. Así, el campo de la guía de alimentación se divide en diferentes frentes de onda que se componen en la apertura de la antena para generar un frente de onda más plano que logra una alta directividad. Además, se imprimen algunas tiras paralelas periódicas después de la apertura de la antena para mejorar la correspondencia de la impedancia con el aire. La antena propuesta tiene un ajuste de impedancia por debajo de -10 dB de 34 a 36 GHz con una ganancia casi constante entre 13 y 15 dBi. Se fabrica un prototipo para validar el diseño de antena propuesto. Como futuras líneas actualmente se está diseñando un *array* en el plano E de 4 antenas base.

**Palabras clave**—Antena de bocina plano H, divisiones de guías, ondas milimétricas, tecnología SIW.

## I. INTRODUCCIÓN

En las comunicaciones móviles, se necesitan nuevos sistemas para la transmisión de velocidades de datos más elevadas. Un cambio en el paradigma actual es el sistema de comunicaciones 5G que tendrá que centrarse en las bandas de frecuencias más altas para resolver los problemas de ancho de banda [1]. Sin embargo, la utilización de estas bandas de frecuencias plantea el problema de la atenuación debida a las pérdidas por propagación. Por lo tanto, las antenas de alta ganancia en bandas de ondas milimétricas son necesarias para superar este último problema. La tecnología *Microstrip* ha sido la tecnología de guiado convencional para el diseño y la fabricación de antenas en el rango de microondas [2]. Sin embargo, cuando la frecuencia de funcionamiento aumenta, la tecnología de *microstrip* sufre pérdidas tales como las pérdidas por radiación. La tecnología de *Substrate-Integrated Waveguide* (SIW) proporciona una solución para reducir las pérdidas por propagación en circuitos planares a frecuencias de onda milimétrica [3]-[5]. En [6]-[8] se presentan algunos ejemplos de diferentes diseños de antenas con alta directividad en esta gama de frecuencias. El último diseño [8] presenta diseños de antenas tipo *endfire* SIW basado principalmente en las sub-aperturas de la antena. El principal inconveniente de estas antenas de apertura es la pérdida de directividad debido al error de frente de fase [9]. Varios trabajos han intentado resolver este problema para mejorar la directividad y ganancia de las antenas SIW de apertura [10]-[16]. Las

soluciones propuestas en la investigación son cargas o la inserción de elementos al final de la apertura, como tiras de metal, orificios para el aire o metasuperficies. Este documento presenta un diseño compacto de antena SIW que aumenta la ganancia de una antena de apertura en frecuencias de ondas milimétricas (banda Ka). El diseño de la antena se basa en la estructura de apertura SIW del plano H, en la que el campo electromagnético se divide por medio de unas vías metálicas interiores integradas en la apertura de la bocina. El documento resumen está organizado de la siguiente manera. En la Sección II se presenta el diseño completo de la antena. La Sección III proporciona la validación experimental del diseño a través de un prototipo. En la sección IV se muestra el planteamiento de *array* en el plano E. Por último, las conclusiones y futuras líneas se exponen en la sección V.

## II. DISEÑO DE LA ANTENA

### A. Configuración de la antena

La configuración geométrica propuesta de la antena SIW compacta en el plano H es mostrada en la Fig. 1. El coaxial de la guía de onda de alimentación se encuentra en el lado derecho de la figura, mientras que la apertura de la antena está en el lado opuesto. En la parte central, se encuentra la estructura de las orificios metalizados internos que implementan las divisiones del frente de onda originadas en la alimentación coaxial.

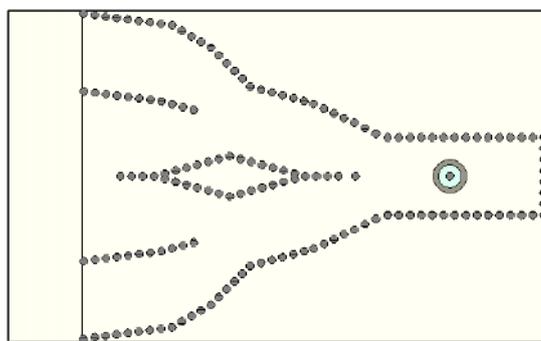


Fig. 1. Vista interior superior de la antena propuesta para el plano H de la SIW. Las dimensiones son 38.9 mm x 23.9 mm.

### B. Diseño y adaptación de la estructura de alimentación

La alimentación de la antena está diseñada como una guía de onda rectangular basada en la tecnología SIW donde un

alimentador coaxial genera el modo fundamental en la guía de onda como se ilustra en la Fig. 2(a). El modo en la guía de onda es TE<sub>10</sub> y para asegurar la excitación del modo fundamental en la frecuencia de trabajo (banda Ka), la altura y el ancho de la guía de onda son 1,524 mm y 5,1 mm, respectivamente. Las placas superior e inferior tienen un espesor de 35 μm y las vías metalizadas que forman la guía de onda tienen una periodicidad de 0,8 mm y un radio de 0,25 mm. Ajustando una distancia adecuada entre la alimentación coaxial y la pared de la vía trasera, es posible obtener una correspondencia adecuada en la banda de frecuencias deseada, como se muestra en la Fig. 2(b).

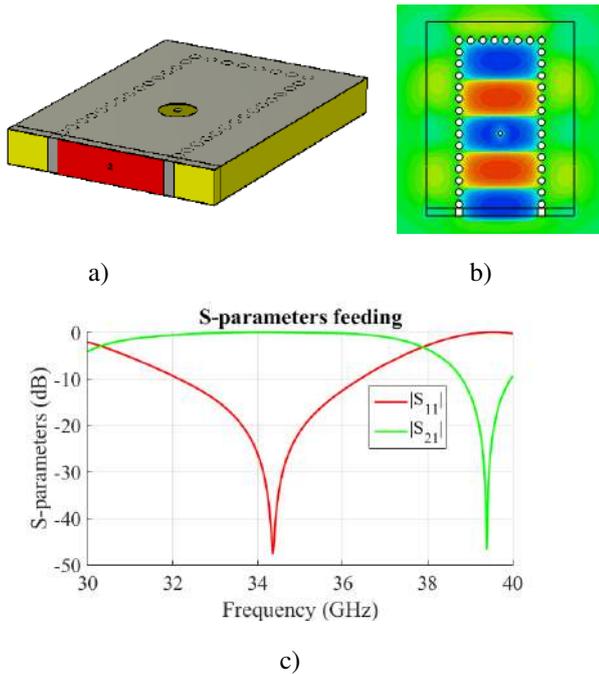


Fig. 2. Alimentación de la guía de onda coaxial SIW: a) Diseño, b) Distribución de campo electrónico, y c) Coeficientes de reflexión y transmisión.

### C. Primera planificación estructura como bocina

Un primer planteamiento para la estructura fue el diseño a partir de una bocina en el plano H a partir de la guía de alimentación anterior. Dentro de la bocina se realizaban orificios metalizados que dividían el frente de onda. En la Fig. 3 se observa la estructura final de esta antena con una primera pared central simulada a partir de los pines que haría de división 1 a 2 y luego otras paredes entre las exteriores y la central que realizarían la división 1 a 4.

El ángulo de las paredes intermedias y su posición de comienzo fueron variables a estudiar en el diseño junto con un gran estudio acerca de cargas o transiciones. El ángulo modificaba el parámetro  $\beta$  que ajustaría la velocidad para compensar las diferencias de fase por la diferencia de caminos de las aperturas. La posición variaba la relación potencia entre las aperturas.

### D. Diseño de la estructura de la guía

La idea anterior fue desechada ya que se tenía un mal control de la fase de las aperturas y no era replicable. Entonces

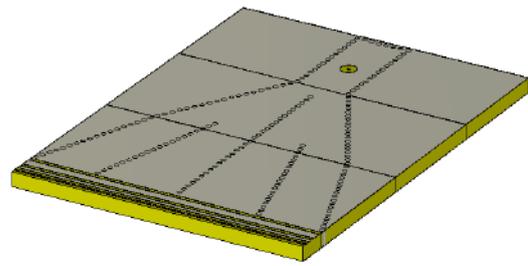


Fig. 3. Antena final con planteamiento de bocina.

se comenzó de nuevo el diseño a partir de la alimentación SIW diseñada (centrada en 35GHz), y en la parte de apertura de la antena se implementa una estructura guiada basada en las divisiones de las guías de onda que se muestra en la Fig. 4(a). Para orientar los frentes de onda hacia la parte frontal de la apertura y también reduciendo la diferencia de fase entre las 4 salidas de la antena, se realiza una unión progresiva de las guías. Otro aspecto relevante es la optimización de los anchos de las guías de onda en la parte final de la apertura de la antena para mantener la misma potencia en las salidas. El ángulo de las hileras de los orificios metalizados en la pared exterior de las guías de onda laterales ajusta la constante de propagación ( $\beta$ ) para conseguir señales de salida en fase. La Fig. 4(b) ilustra la distribución del campo E a lo largo de la antena propuesta y revela la fase casi constante y la amplitud en las salidas de la apertura de la antena.

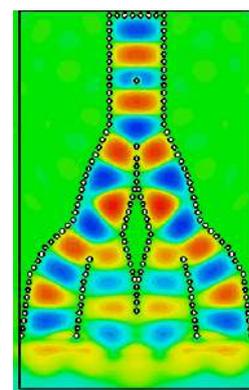
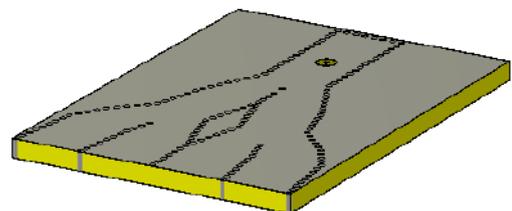


Fig. 4. Diseño optimizado de la estructura de la guía: a) modelo de estructura, b) distribución de campo a 35 GHz.

### E. Interfaz sustrato-aire y diseño completo

El elemento de adaptación de la impedancia desde la apertura de la antena al aire se describe en esta subsección.

Por medio de una losa dieléctrica del mismo sustrato con tiras metálicas (en las partes superior e inferior) delante de la apertura de la antenna, se puede mejorar la adaptación de la impedancia[17]. La transición se compone de 3 pares horizontales de bandas metálicas y bandas vacías. Cada par de bandas tiene una longitud de un cuarto de longitud de onda, pero la relación entre la banda metálica y la longitud de la banda de aire se varía progresivamente acorde a cierto patrón: desde el 80% de la banda metálica hasta el 20% de la banda de aire al 20 % de la banda metálica y al 80% de la banda de aire. El par de bandas del medio es un 50% metálico y 50% corresponde a la banda de aire. La Fig. 5(a) muestra el diseño de la antenna con la transición al aire. El coeficiente de reflexión para el diseño final de la antenna se muestra en la Fig. 5(b), que produce una impedancia y una anchura de 34 a 36 GHz. Además, el patrón de radiación 3D es ilustrado para mostrar el rendimiento de la directiva de la antenna.

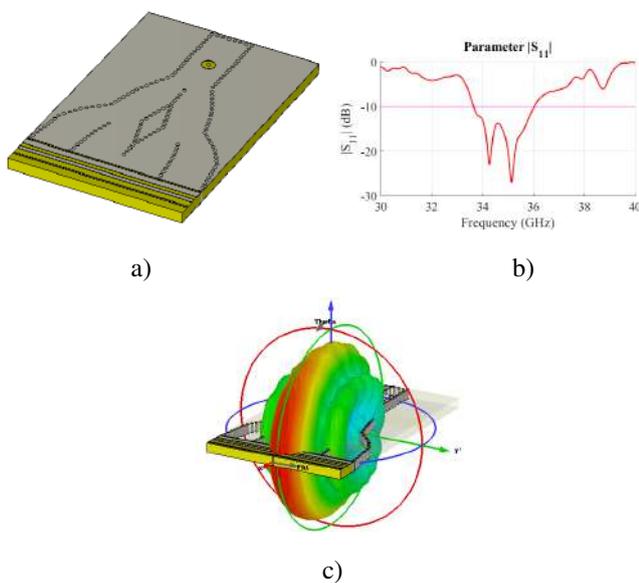


Fig. 5. Diseño final de la antenna: a) modelo de estructura, b) parámetro  $|S_{11}|$ , y c) patrón de radiación (3D) a 35 GHz.

### III. PROTOTIPOS Y MEDIDAS

Una vez terminado el diseño, el prototipo es manufacturado y medido. La permitividad relativa fabricada ( $r$ ) del sustrato es de 3,38. Las Figs. 6(a) y 6(b) muestran el prototipo y el conector de 2,92 mm adjunto para alimentar la antenna. El coeficiente de reflexión medido se ilustra en la Fig.6(c). Una pequeña desviación para bajar frecuencias, esto puede deberse a las tolerancias en los  $\epsilon_r$  manufacturados. Las mediciones de radiación de la antenna se realizan en la cámara anecoica de la Universidad de Granada (UGR). En las Figs. 7(a) y 7(b) se muestran la configuración de la medición. Se muestran los diagramas de radiación del plano H a diferentes frecuencias en la Fig. 7(c)-7(e). Los diagramas de radiación direccionales se observan para la banda de frecuencia de funcionamiento de la antenna fabricada. En frecuencias más altas, el lóbulo principal es más estrecho y la directividad tiende a aumentar. Finalmente, la ganancia de la antenna es medida y se ilustra en la Fig. 7(f). Aunque la impedancia es de 33 a 35.5 en el

prototipo, se obtiene una mayor ganancia de 34 a 36 porque el frente de onda en la apertura de la antenna está optimizada para este rango de frecuencia. Además, la razón por la que la ganancia comienza a disminuir de 35.5 GHz es el desajuste de impedancia, ya que el desajuste medido se degrada en comparación con los resultados simulados (frecuencia turno). Sin embargo, se consigue una ganancia estable superior a 13 dBi de 34 o 36 GHz.

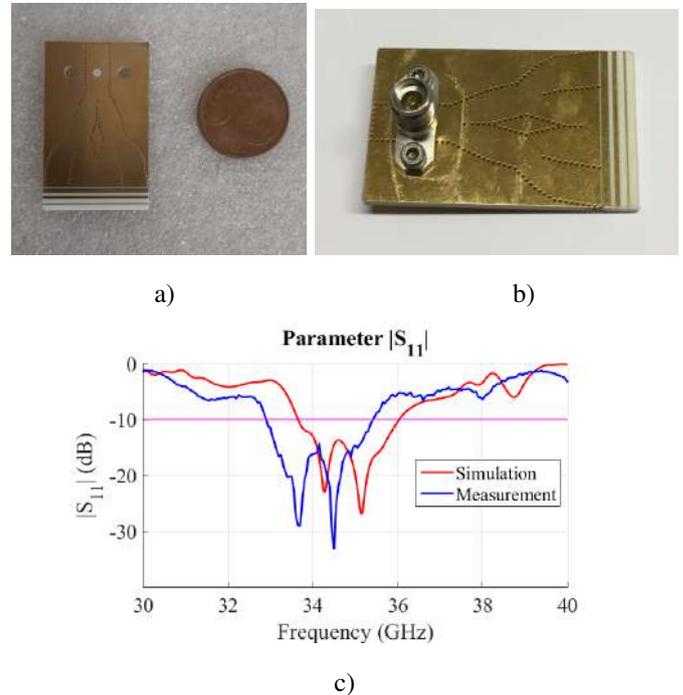


Fig. 6. Prototipo de antenna: a) vista superior, b) vista inferior y c) comparativa del parámetro  $|S_{11}|$ .

### IV. ARRAY PLANO E

En esta sección se explicará el proceso seguido para un primer planteamiento hacia una de las líneas futuras, una agrupación de la antenna base ya construida. Ya que la antenna base puede ser considerada como una agrupación de 4 aperturas (la fase y potencia que llegan a las aperturas exteriores es algo diferente a las centrales), podría formarse un *array* en dos dimensiones (4x4) simplemente colocando 4 antenas base apiladas.

La problemática de la idea sería la alimentación coaxial que para que consiga alimentar las 4 antenas de una forma similar (fase parecida e igual potencia) se necesitan de placas metálicas con un orificio del mismo material que la guía SIW. Esto es así para evitar pérdidas por cambios de impedancia y además obtener gracias al espesor de las placas la distancia inter-antena que se necesita para una fase progresiva nula. En este caso en la Fig. 8(a) se comprueba que con un espesor de 4mm se consigue el efecto (interdistancia de 5.5mm, porque las antenas tienen espesor de 1.5mm). En la Fig. 8(b) se observa el aumento de la directividad deseado.

### V. CONCLUSIÓN

Este proyecto propone un diseño optimizado de una antenna en tecnología de integración de guía de onda en sustrato

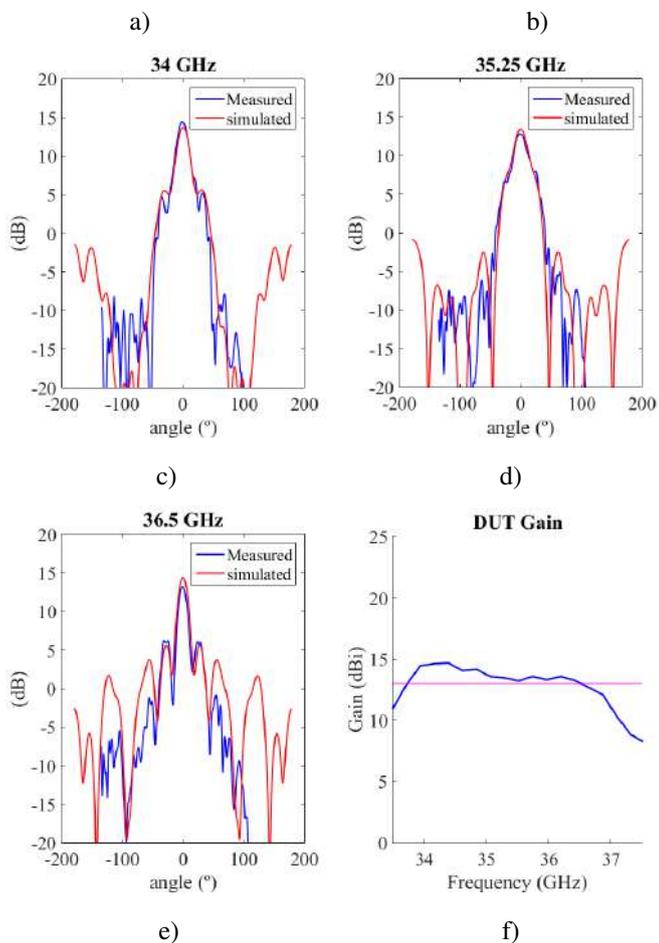
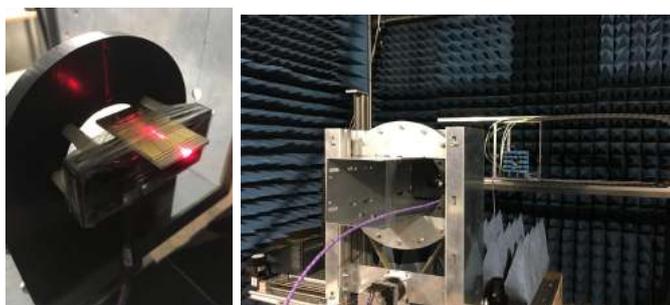


Fig. 7. Medidas de la antena: a) proceso de alineación de la antena, b) configuración de la medición del diagrama, c) diagrama de radiación del plano H de 34 GHz, d) diagrama de radiación del plano H de 35,25 GHz, e) diagrama de radiación del plano H de 36,5 GHz, y f) ganancia.

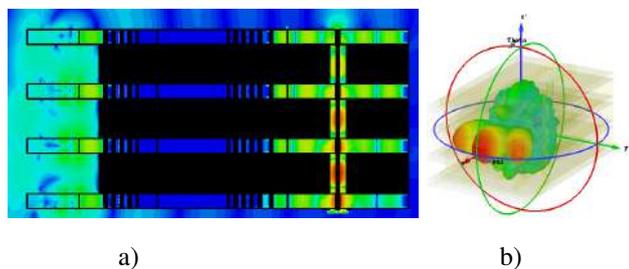


Fig. 8. Array plano E: a) Diseño, y b) Diagrama de radiación a 35 GHz.

(SIW) basada en la estructura de una antena de apertura en el plano H. La estructura de la antena está diseñada a través

de las divisiones de las guías de onda unidas (economizando el número de las vías en las hileras) y enfocando el campo en la apertura para mejorar la directividad de la antena. De acuerdo con las medidas realizadas, existe un buen acuerdo entre las simulaciones y las medidas del prototipo.

La única diferencia entre la simulación y los resultados experimentales es un desplazamiento de frecuencia de 0,5 GHz de la banda de frecuencias de funcionamiento debido a las tolerancias de fabricación. El rendimiento de radiación de la antena propuesta es alto debido a los frentes de onda planos reforzados en 4 salidas de guía de onda que forman la apertura de la antena. Se ha logrado una ganancia media estable en la banda de frecuencias de funcionamiento de alrededor 13dBi.

Como futuras líneas abiertas en el desarrollo y mejora de esta antena se está diseñando un *array* en el plano E que mejoraría la ganancia total de la antena a base de mejorar la directividad en el plano no previsto de la antena base. Otra línea en actual desarrollo es la incorporación la planificación de esta antena como una división de guías SIW. Ambas dentro del marco de la cámara anecoica del CITIC.

#### AGRADECIMIENTOS

Este trabajo ha sido parcialmente apoyado por el TIN2016-75097-P, RTI2018-102002-A-I00 y EQC2018-004988-P proyectos del Programa Nacional de Investigación y Desarrollo de España, Innovación, así como el proyecto B-TIC-402-UGR18 de la Junta de Andalucía.

#### REFERENCIAS

- [1] T. S. Rappaport, S. Sun, R. Mayzus., H. Zhao, Y. Azar, K. Wang. "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!", in *IEEE Access.*, vol. 1, pp. 335 - 349, May 2013.
- [2] D. M. Pozar and D. H. Schaubert. *Microstrip Antennas: The Analysis and Design of Microstrip Antennas and Arrays*, Wiley, 1995.
- [3] D. Deslandes and K. Wu, "Integrated microstrip and rectangular waveguide in planar form", in *IEEE Microw. Wireless Compon. Lett.*, vol. 11, no 2, pp. 68-70, Feb. 2001.
- [4] F. Xu and K. Wu, "Guided-wave and leakage characteristics of substrate integrated waveguide", in *IEEE Trans. Microw. Theory Techn.*, vol 53, no. 1, pp. 66-73, Jan. 2005.
- [5] K. Wu, Y. J. Cheng, T. Djeraji and W. Hong, "Substrate-Integrated Millimeter-Wave and Terahertz Antenna Technology", in *Proceedings of the IEEE*, vol. 100, no. 7, pp. 2219-2232, July 2012.
- [6] Y. J. Cheng et al., "Substrate Integrated Waveguide (SIW) Rotman Lens and Its Ka-Band Multibeam Array Antenna Applications", in *IEEE Transactions on Antennas and Propagation*, vol. 56, no. 8, pp. 2504-2513, Aug. 2008.
- [7] L. Wang, J. L. Gomez-Tornero and O. Quevedo-Teruel, "Substrate Integrated Waveguide Leaky-Wave Antenna With Wide Bandwidth via Prism Coupling", in *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 6, pp. 3110-3118, June 2018.
- [8] Y. Cao, Y. Cai, L. Wang, Z. Qian and L. Zhu, "A Review of Substrate Integrated Waveguide End-Fire Antennas", in *IEEE Access*, vol. 6, pp. 66243-66253, 2018.
- [9] W. Stutzman and G. Thiele. *Antenna Theory and Design*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.
- [10] L. Wang, Esquiús-Morote H. Qi, X. Yin, and J. R. Mosig. "Phase Corrected H-Plane Horn Antenna Gap SIW Technology", in *IEEE Trans. Antennas Propag.*, vol. 65, no. 1, Jan. 2017
- [11] H. Wang, Da-Gang Fang, B. Zhang, and Wen-Quan Che, "Dielectric Loaded Substrate Integrated Waveguide (SIW) H-Plane Horn Antennas", in *IEEE Trans. Antennas and Propag.*, vol. 58, no. 3, March 2010.
- [12] H. Jamshidi-Zarmehri and M. H. Neshati, "Design and Development of High-Gain SIW H-Plane Horn Antenna Loaded With Waveguide, Dipole Array, and Reflector Nails Using Thin Substrate", in *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 4, pp. 2813-2818, April 2019.
- [13] Y. Luo and J. Bornemann, "Substrate Integrated Waveguide Horn Antenna on Thin Substrate With Back-Lobe Suppression and Its Application to Arrays", in *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 2622-2625, 2017.

- [14] N. Bayat-Makou and A. A. Kishk, "Substrate Integrated Horn Antenna With Uniform Aperture Distribution", in IEEE Transactions on Antennas and Propagation, vol. 65, no. 2, pp. 514-520, Feb. 2017.
- [15] Y. Cai, Y. Zhang, Z. Qian, W. Cao and L. Wang, "Design of Compact Air-Vias-Perforated SIW Horn Antenna With Partially Detached Broad Walls", in IEEE Transactions on Antennas and Propagation, vol. 64, no. 6, pp. 2100-2107, June 2016.
- [16] L. Gong, K. Y. Chan, and R. Ramer, "Substrate integrated waveguideHplane horn antenna with improved front-to-back ratio and reduced sidelobe level", in IEEE Antennas Wireless Propag. Lett., vol. 15, pp. 1835-1838, 2016.
- [17] L. Wang, M. Garcia-Vigueras, M. Alvarez-Folgueiras, and J. R. Mosig. "Wideband H-plane dielectric horn antenna", in IET Microw. Antennas Propag., vol. 11, iss. 12, pp. 1695-1701, 2017



**Cleofás Segura Gómez** nació el 15 de marzo de 1997 en Almería, estudió educación secundaria en el IES Bahía de Almería y terminó el graduado de Ingenierías de Tecnologías de las Telecomunicaciones (mención en Sistemas de Telecomunicación) en 2019 por la Universidad de Granada.

Actualmente (2019) se encuentra cursando el máster profesional en Ingeniería de Telecomunicación en la Universidad de Granada.



# PROTOTIPO DE UNA ESTACIÓN BASE 4G USANDO OPEN AIR INTERFACE

Autor: Francisco García Espigares, e-mail: franx24@correo.ugr.es

Tutores: Jorge Navarro Ortiz, e-mail: jorgenavarro@ugr.es

José Carlos Segura Luna, e-mail: segura@ugr.es

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—En los últimos años y de forma más acelerada, el mundo de las TIC (*Tecnología de la Información y la Comunicación*) está llevando a cabo una continua transformación a través de múltiples aspectos, como la aparición de nuevas tecnologías o la implantación de innovaciones en los sistemas *software* y *hardware*, llevando consigo oportunidades que posibilitan el nacimiento de nuevos mercados.

Por estos motivos, el objetivo principal de este proyecto consiste en la puesta en marcha de una celda *Long Term Evolution* utilizando para ello el *software* libre *Open Air Interface*, un dispositivo *Universal Software Radio Peripheral*, y un PC comercial. Esto conlleva la correcta instalación y configuración de la red de acceso radio o *Evolved Universal Terrestrial Radio Access Network*, además de la red troncal o *Evolved Packet Core*. Dicho proyecto concluirá con la evaluación de dicha plataforma, tanto desde el punto de vista de las señales radio como de la señalización empleada.

**Palabras clave**—*Long Term Evolution*, *Open Air Interface*, *Open Air Interface System Emulation*, *Sysmocom*, *Universal Software Defined Radio*, *VMware*, *Universal Subscriber Identity Module*, *4G*.

## I. INTRODUCCIÓN

CON el lanzamiento en los últimos años del despliegue de redes de cuarta generación, el aumento en la demanda del acceso a servicios de Internet por parte de usuarios de teléfonos móviles, y la transformación de una sociedad cada vez más involucrada con el campo de las Tecnologías de la Información y la Comunicación, surge la necesidad de crear soluciones sencillas, abiertas y que se adapten a cualquier situación. El presente proyecto trata de realizar un estudio de una red celular LTE, basada en el estándar del 3GPP, con la plataforma *Open Air Interface*. Así, se pretende desplegar una pequeña red celular propia, estable y que cumpla con los requisitos de la normativa. Esto permitirá la simplificación en la utilización de recursos y, con ello, reducir los costes que puede llevar consigo el despliegue de dicha red.

Gracias a la reducción de costes de los equipos radio definidos por *software*, el aumento de prestaciones de los PCs y la virtualización de equipos, es posible diseñar un servicio de comunicaciones tan complejo como puede ser una red móvil de cuarta generación mediante un equipamiento de bajo coste.

## II. RESUMEN DEL ESTÁNDAR LTE

Este apartado será un breve resumen de las características principales del estándar LTE, de forma que el lector que no

esté familiarizado con él pueda comprender los conceptos de los siguientes apartados.

LTE presenta una arquitectura simplificada en comparación con tecnologías anteriores, ya que está separada en dos bloques bien diferenciados: el núcleo de red o CN, y la red de acceso radio o RAN. Los nombres con los que se les conoce comúnmente a dichos bloques en LTE son EPC, para el núcleo de red, y E-UTRAN para la red de acceso radio.

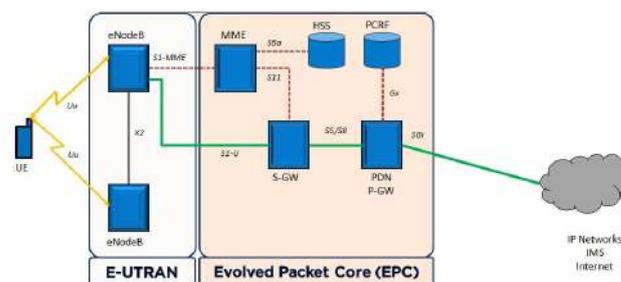


Fig. 1. Arquitectura LTE.

El objetivo de diseño para la arquitectura LTE es minimizar el número de nodos, por lo que ha conducido a la implantación de un nodo más complejo que el NodeB de UMTS, llamado *evolvedNodeB*. Un *eNodeB* tiene como función principal la gestión de los recursos radio y la conexión de los terminales móviles a la red.

### A. EPC

En esta sección se pretende dar una breve visión del EPC (*Evolved Packet Core*), de las entidades que lo componen y sus funcionalidades. El núcleo de red, aparte de realizar funciones propias del Core Network como son la facturación, *roaming*, interconexión con redes externas, etc. incorpora las siguientes funciones:

- De transporte: soportan la transmisión de información de tráfico y señalización.
- De inteligencia: llevan a cabo aspectos de encaminamiento, gestión de movilidad, control de los servicios ofertados la calidad de éstos.

La arquitectura del núcleo de red de LTE implementa un único dominio de paquetes basado en la arquitectura TCP/IP,

dando lugar a la solución llamada *all IP*. Las principales entidades del núcleo de red en LTE son:

1) *Home Subscriber Server (HSS)*: El HSS es una base de datos central que presta servicio a todas las entidades de control que gestionan sesiones y llamadas.

2) *Equipment Identity Register (EIR)*: El EIR es la entidad lógica que se encarga de almacenar las identidades internacionales de los equipos móviles (IMEI) en distintas listas.

3) *Mobility Management Entity (MME)*: El MME es el encargado del plano de control dentro del núcleo de red. Las funcionalidades que esta entidad posee son: señalización y seguridad, movilidad, seguimiento, etc.

4) *Serving Gateway (S-GW)*: El nodo S-GW es el principal punto del EPC para el enrutamiento y encaminamiento de paquetes.

5) *Packet Data Network Gateway (PDN-GW)*: El nodo PDN-GW es responsable de actuar como un punto fijo de la movilidad entre tecnologías 3GPP y no 3GPP.

6) *Policy and Charging Rules Function (PCRF)*: La entidad PCRF es responsable de la política y control de cargos, decide la forma de manejar los servicios en términos de QoS.

### B. E-UTRAN

La parte E-UTRAN tiene una estructura de red plana, por lo que los eNB pueden comunicarse directamente entre sí, sin tener que realizar la señalización y el envío de mensajes a través de un nodo central.

1) *evolved Node Base (eNB)*: El eNB se encarga de la transmisión de los paquetes IP hacia y desde los equipos del usuario hasta el núcleo de red EPC, en los que se envía la información de dicho usuario y los mensajes de señalización.

## III. OPEN AIR INTERFACE

Para la realización de este proyecto, hemos apostado por el desarrollo del *software* libre basado en el estándar 3GPP. El *software* escogido es *Open Air Interface (OAI)*, el cual está escrito en lenguaje C para Linux, proporcionando un ecosistema de código libre para los protocolos del núcleo de red o EPC, y para la red de acceso o E-UTRAN de los sistemas celulares 3GPP, basados en los estándares de la Rel.10 para LTE.

El *software* de OAI nos ofrece un amplio abanico de posibilidades a la hora de montar nuestro escenario, proporcionando un alto grado de flexibilidad, permitiendo adaptarse a los recursos *hardware* de los que dispongamos. En el presente proyecto realizaremos el análisis de estos dos escenarios, ya que no disponemos en el laboratorio de eNB y EPC comerciales.

- OAI EPC + OAISIM
- OAI EPC + OAI eNB + UE (comercial)

## IV. INSTALACIÓN Y CONFIGURACIÓN

### A. EPC+OAISIM

En primer lugar, hemos realizado la simulación del escenario real en máquinas virtuales. El diseño de las máquinas virtuales dentro de nuestro equipo portátil se puede apreciar en la figura 2.

Una vez creadas las dos máquinas virtuales realizamos la instalación de *Ubuntu Linux 14.04 LTS(64 bits)*. A continuación, se modifica la configuración de red de ambas máquinas para obtener el diseño visto en la figura 2.

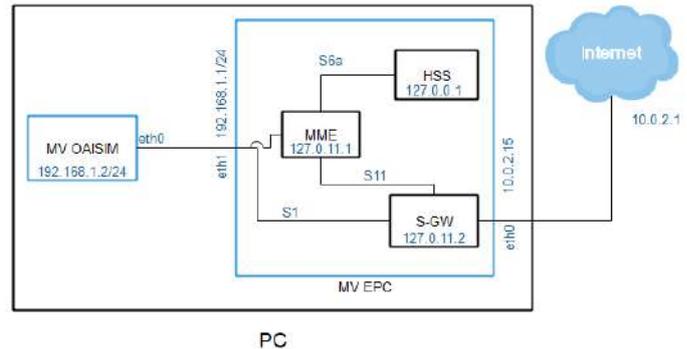


Fig. 2. Estructura lógica de la conexión de las MV.

A continuación, hemos realizado la instalación de OAI CN y de OAI SIM en máquinas virtuales diferentes, para posteriormente trasladar dicha configuración a los equipos reales situados en el laboratorio. Para realizar la instalación de las distintas entidades en cada máquina virtual, procedemos a descargar los códigos fuente desde gitlab de OAI [3]. Acto seguido, nos dirigiremos al directorio donde hemos descargado el repositorio y se encuentran los códigos fuente, para realizar la instalación y compilación de las entidades MME, HSS y S-GW.

1) *Configuración del HSS*: En primer lugar, lo que debemos hacer para comenzar con la configuración del CN (en el equipo “epc”) es la modificación de la base de datos que utilizará el HSS. En la tabla *mmeidentity* incluiremos el nombre de los equipos involucrados y el dominio de éstos.

Una vez concluida la modificación e introducción de datos en la base de datos vamos a proceder a modificar el archivo de configuración del equipo virtual HSS. El cual nos permitirá configurar los parámetros para la conexión a la base de datos. Otro archivo a modificar antes de poner en funcionamiento el HSS es el *hss\_fd.conf*. En el que estableceremos la configuración para la conexión con la entidad MME de nuestra red LTE, como puede ser el dominio, la dirección IP del MME, las rutas de los certificados TLS, etc. La posterior acción a realizar una vez concluida la modificación de los archivos anteriores es crear las claves y certificados de autenticación.

2) *Configuración del MME*: A continuación, una vez en funcionamiento el HSS, seguimos con la modificación del archivo de configuración del MME, el *mme.conf*. En la simulación no modificaremos los parámetros de la lista GUMMEI, ni de la lista TAI, por lo que permanecerán por defecto. Los parámetros que modificaremos de acuerdo con la topología mostrada en la figura 2, serán las direcciones IP, el nombre de dominio, los nombres de las distintas entidades que componen la red LTE. Otro archivo que modificaremos es el denominado *mme\_fd.conf* el cual nos permite modificar los parámetros para la conexión entre el MME y el HSS a

través del interfaz S6A, que como sabemos nos proporciona el soporte para la movilidad. La posterior acción a realizar una vez concluida la modificación de los archivos anteriores es crear las claves y certificados de autenticación. Se obtienen del mismo modo que para el HSS. Finalmente, una vez que tenemos terminada la configuración de los archivos anteriores, procedemos a poner en funcionamiento el MME.

3) *Configuración del S-GW*: Por otra parte y para terminar la configuración del equipo que actúa como *Core Network* realizaremos la modificación del archivo de configuración para el S-GW, denominado `spgw.conf`. En este archivo, debemos de establecer la dirección IP del MME, la dirección IP del equipo que nos dará el acceso al exterior, y el conjunto de direcciones IP que estarán disponibles para los usuarios de nuestra red LTE.

Finalmente, una vez que hemos concluido de modificar el archivo de configuración, debemos de compilar y lanzar el S-GW.

### B. Configuración de OASIM

En este apartado procederemos con la configuración de OASIM, en la máquina virtual OASIM, tal y como se muestra en la figura 2. Además, hemos asumido que se ha realizado la instalación de OASIM anteriormente.

En primer lugar, lo que debemos hacer es modificar el archivo de configuración, `enb.band7.generic.oasim.local_mme.conf`.

Entre los parámetros que hemos modificado en este archivo son: la identificación de celda y del eNodeB, los códigos de área (TAC, MCC y MNC), además de parámetros físicos como la banda utilizada, tipo de modulación, frecuencia central para el enlace ascendente y descendente, los parámetros de configuración de red para el MME, etc.

### C. OAI EPC + OAI eNB(USRP B210) + UE

En esta sección trasladaremos la configuración realizada en la emulación al escenario real lo primero que debemos realizar es grabar los datos de las tarjetas USIM que utilizaremos para poder conectarnos con dispositivos móviles comerciales a nuestra red LTE.

1) *Configurando el OAI CN*: En este apartado configuraremos todas las entidades que componen el EPC, como ya vimos anteriormente en el apartado de *Simulación*.

En primer lugar comenzaremos con la introducción de datos en la base de datos que utilizará el HSS para la autenticación de los UE, como es el IMSI, ICCID, ACC, PIN1, Ki, OP, OPC, etc. La instalación y configuración de todas las entidades que componen el *core network* para la implementación del escenario real se llevan a cabo de la misma forma que en el escenario anterior, en el que se ha realizado la emulación.

2) *Configuración del OAI eNB*: La entidad que actúa como eNodeB es un *hardware* que utilizaremos como estación base es el USRP B210, el cual irá conectado mediante un USB 3.0 a nuestro PC. Para realizar por primera y única vez la instalación de unos paquetes que son necesarios para que OAI eNB funcione correctamente, y una vez terminada la

instalación del *software* es necesario realizar la modificación del archivo de configuración que es utilizado para el USRP B210, de la misma forma que en el caso de la emulación.

3) *Configuración del UE y USIM*: Además, de la instalación y configuración del HSS, MME, S-GW debemos de realizar la configuración en cada móvil comercial que emplearemos. Para ello debemos crear correctamente un APN, con sus respectivos valores.

Otro de los procedimientos que debemos llevar a cabo es la programación de las tarjetas USIM que nos permitirán acceder a nuestra red LTE.

## V. ANÁLISIS

En primer lugar, se ha realizado un análisis de los mensajes de señalización que se llevan a cabo en el escenario de emulación. Posteriormente, realizamos un análisis del escenario real que se ha creado a partir de un PC de sobremesa, un USRP B210, y dos teléfonos móviles realizando el rol de UE's.

### A. EPC + OASIM

En la figura 3 se ha obtenido el flujo de mensajes intercambiados entre las máquinas virtuales OASIM y EPC. Se han capturado con Wireshark y los hemos filtrado con el protocolo S1-AP para quedarnos exclusivamente con los mensajes de señalización.

La primera secuencia de mensajes que nos encontramos son los mostrados en la figura 3, que se envían desde el eNB que inicia la conexión hacia el MME, y viceversa para establecer la conexión del eNB a la red LTE. En el primer mensaje

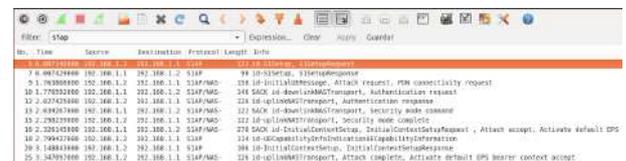


Fig. 3. Señalización EPC + OASIM.

de la estación base, el *S1 Setup Request*, se envía diversa información desde la identidad global del eNB como es el MCC y el MNC, hasta el propio nombre de la estación base.

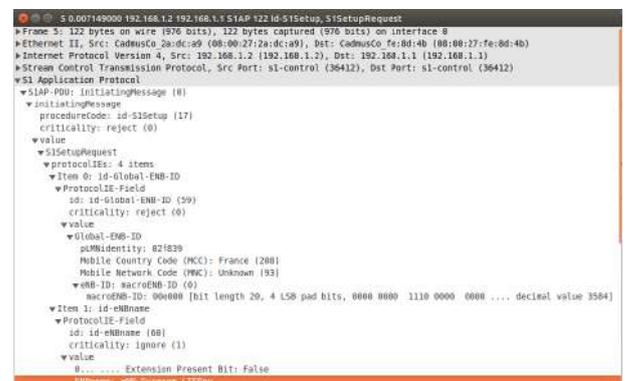


Fig. 4. Parte del mensaje S1 Setup Request.

Debido a la extensión no podemos explicar la función de cada uno de estos mensajes, pero sí está disponible en el proyecto previamente redactado.

**B. OAI EPC + OAI eNB + UE**

En esta sección, vamos a realizar un análisis del escenario real mostrado en la figura 5. En el cual, como podemos apreciar, tenemos instalado y funcionando el OAI EPC, OAI eNB y dos teléfonos móviles que tendrán el papel de UE. A continuación, se detalla un breve resumen con las características físicas de los distintos escenarios evaluados.

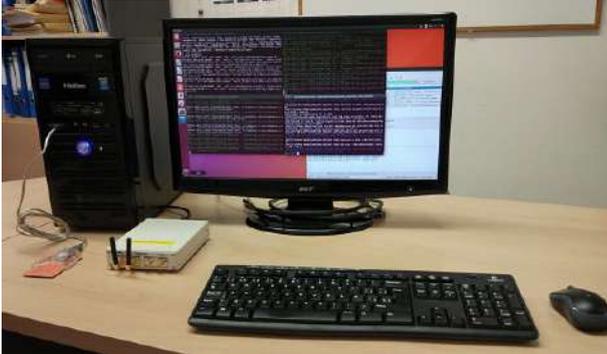


Fig. 5. Aspecto del escenario real.

terísticas físicas de los distintos escenarios evaluados.

- Parte radio.
  - Frecuencia central:
    - \* UL: 2,565 GHz
    - \* DL: 2,685 GHz
  - N° de RB: 25, 50, 100
  - Ancho de banda:
    - \* UL: 5 MHz, 10 MHz, 20 MHz
    - \* DL: 5 MHz, 10 MHz, 20 MHz

1) *Parte radio:* En este apartado evaluaremos la capa física, examinando el ancho de banda, ganancias, slot temporales, la PSD y el espectrograma, tanto en el enlace descendente como en el ascendente. Para ello hemos llevado a cabo varias veces la misma prueba, en la que mediante la aplicación “SpeedAnalytics” se ha realizado un test de velocidad utilizando los teléfonos móviles disponibles para el proyecto. El análisis de la parte radio se ha realizado mediante el analizador de espectros AGILENT N9010A, y con scripts del *toolbox* LTE de MATLAB.

*Analizador de espectros*

Mientras se realizaban los test de velocidad, con el analizador de espectros conseguimos capturar el ancho de banda que se utilizaba para el enlace ascendente, y así poder realizar varias medidas.

Una vez obtenidos ambos límites, podemos comprobar que OAI eNB para el enlace descendente utiliza 4.52 MHz. Para DL sí cumple el estándar establecido del 3GPP para el ancho de banda ocupada para la transmisión, teniendo en cuenta que debe tener 5 MHz de anchura nominal. Si realizamos los cálculos para comprobar que el ancho de banda es correcto, tenemos:  $25RB * 180kHz = 4,5MHz$ , si le añadimos los 250 kHz de banda de guarda tanto en el límite inferior como en el superior, tenemos los 5 MHz de ancho de banda nominal, tal y como se especifica en el estándar. Por lo que el resultado obtenido es correcto.

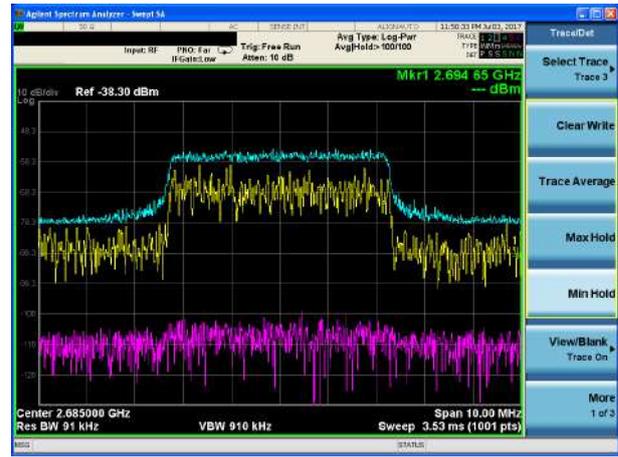


Fig. 6. Máximos y mínimos de la señal el frecuencia.

Estableciendo un ancho de banda de 5 MHz para el canal descendente, en la figura 6 observamos los máximos (color azul) y mínimos (color rosa) de las frecuencias para cada una de las portadoras que componen el canal.

El analizador de espectros AGILENT N9010A tiene instalado un paquete básico (versión educación) para la demodulación de señales LTE. Debemos indicarle la frecuencia en la que está transmitiendo el eNB, y el canal que nos interesa analizar. En la figura 7 se puede apreciar el análisis que

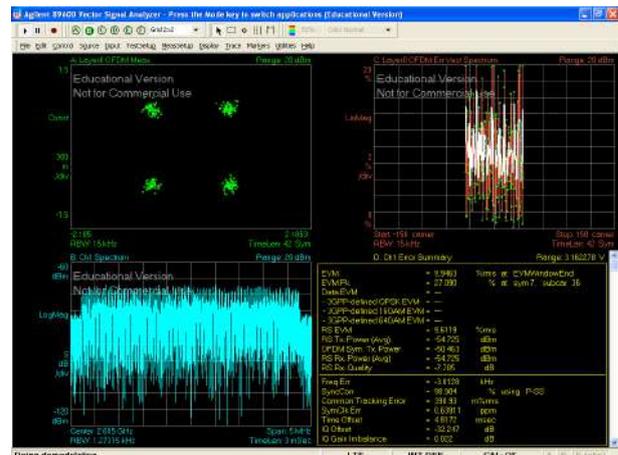


Fig. 7. Análisis del canal PBCH.

realiza para el canal PBCH. En la esquina superior izquierda se encuentra la constelación QPSK que utiliza dicho canal; en el lado opuesto se encuentra el EVM para los símbolos OFDM transmitidos; en la parte inferior izquierda nos encontramos con el espectro de nuestro canal de transmisión (en este caso estamos transmitiendo con 25 RB, que hacen un total de ancho de banda de 5 MHz); y finalmente en la esquina inferior derecha nos muestra un resumen con las características del canal seleccionado, en el que podemos ver entre otras cosas el EVM, la potencia de transmisión para el símbolo OFDM, o el error de frecuencia.

*PSD y espectrograma mediante MATLAB*

En este apartado se ha realizado la captura de la señal LTE de nuestra estación base utilizando Simulink utilizando el paquete *espectro\_lte2.slx* para posteriormente procesarla con un

paquete de MATLAB llamado *LTE System Toolbox™*, el cual puede utilizarse para sincronizar, demodular, y decodificar una señal de un eNB en tiempo real.

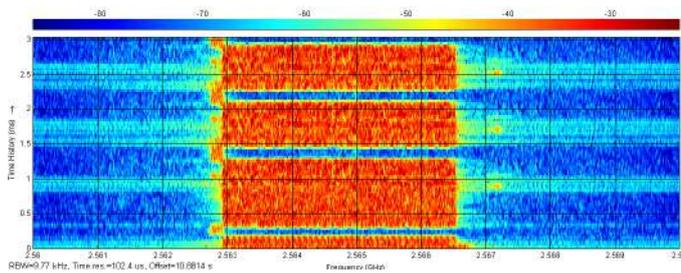


Fig. 8. Espectrograma del intercambio de información en UL.

La figura 8 es el espectrograma una vez que se ha conectado el UE a la red móvil, y se produce el intercambio de información entre el UE y el OAI eNB. Podemos observar diferentes *slots* temporales con la duración establecida por el estándar de 0.5 ms. Estas ráfagas de información pueden ser a causa de actualización de información de la red, solicitudes por parte de la red al UE, o el caso contrario, actualizaciones de aplicaciones por parte del UE, etc.

Además, en la figura 8 observamos que solamente se utilizan 3,5 MHz de los 4,5 MHz disponibles para el ancho de banda de 5 MHz. Esto puede deberse a que el UE utilizado es incapaz de soportar todo el ancho de banda.

Otra característica que podemos observar es en la parte superior del espectrograma, los *slots* de 0.5 ms, mientras que en la parte inferior se ha capturado una subtrama de 1 ms. Ésta es una de las 10 subtramas que forman una trama de LTE. La separación temporal sabemos que está establecida en 0.21ms, que debe ser menor que el tiempo de coherencia del canal, por lo que también comprobamos que se ajusta correctamente a las especificaciones.

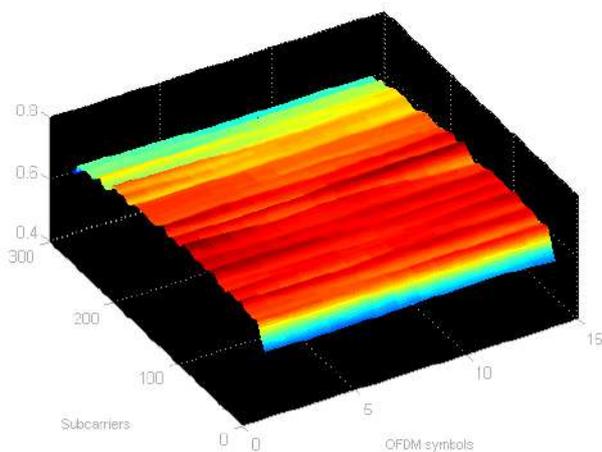


Fig. 9. Respuesta en magnitud del canal para un BW de 5MHz.

A continuación, realizaremos el análisis de la señal capturada a partir de Simulink con el paquete LTETolbox. Una de las imágenes que nos muestra el paquete LTETolbox de MATLAB es la respuesta en magnitud del canal. Como vemos

en la figura 9, para un ancho de banda de 5MHz, LTE establece un total del 301 portadoras. Además, en esta imagen podemos observar la magnitud en el tiempo de 15 símbolos OFDM.

En la figura 10, recoge el espectro obtenido. En este podemos ver el máximo, mínimo y el valor medio del espectrograma de la señal. Vemos que oscila en el rango comprendido entre -30 dBm y -80 dBm.

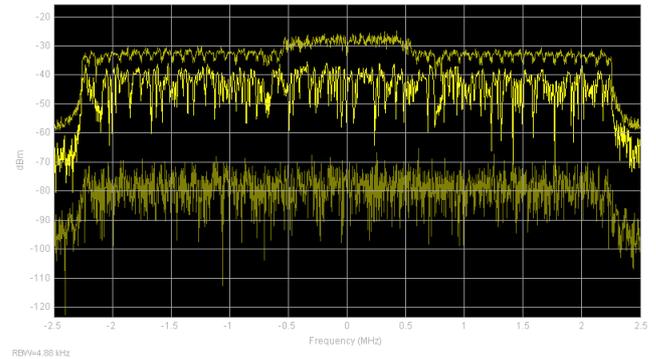


Fig. 10. Espectro del canal para un BW de 5MHz.

En la figura 11, nos muestra la constelación QPSK para el canal decodificado correspondiente al canal PDSCH. Para dicho canal se han estimado los siguientes valores RMS EVM: 6.352%, PDSCH Peak EVM: 15.591%

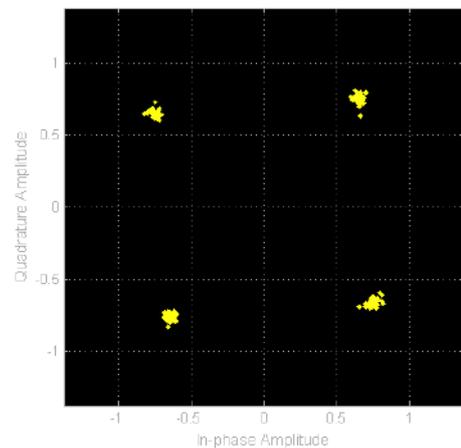


Fig. 11. Constelación del canal para un BW de 5MHz.

La figura 12, se corresponde con la correlación para identificar el principio de las tramas. Se correla la señal recibida con la señal de sincronización de trama (PSS, que es una secuencia de bits). Es necesario la sincronización de trama antes que la demodulación de los bits. A partir de los bits de la PSS se generan los símbolos OFDM, y esa es la señal que se correla con la señal recibida (aún no está demodulada, no se puede correlar a nivel de bits).

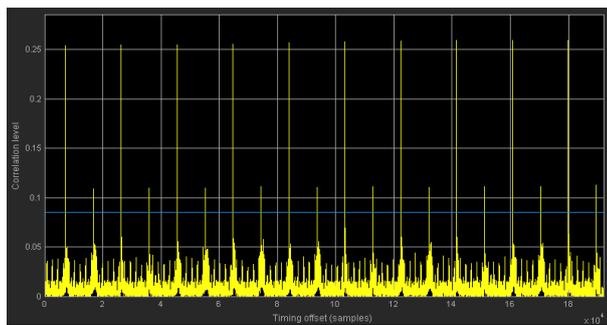


Fig. 12. Correlación del canal para un BW de 5MHz.

### C. Pruebas de velocidad

Se han realizado pruebas de velocidad con dos teléfonos móviles. Para ello hemos descargado la aplicación gratuita "SpeedAnalytics". En una primera prueba hemos realizado la conexión de cada móvil de manera independiente, y de forma que solamente esté un dispositivo conectado a nuestra red 4G en un instante dado. En los test realizados con el terminal Xiaomi, obtenemos en ambos casos una velocidad media en el enlace descendente entorno a 14,5 Mbps; por otro lado, en el enlace ascendente observamos una variación considerable del primer test (8,24 Mbps) respecto del segundo; en éste último la velocidad media es de 14,95 Mbps con lo que tenemos una simetría en el ancho de banda para los dos enlaces. Otro parámetro que obtenemos con esta prueba es el retardo del ping, que eleva su valor respecto del primer test. Repitiendo el procedimiento anterior obtenemos los test para el dispositivo móvil BQ. En el primer test se ha obtenido una velocidad media más baja que en el segundo, tanto para el enlace ascendente como para el descendente. En el primero se ha obtenido una velocidad media en el enlace UL de 8,47 Mbps, mientras que en DL una velocidad de 15,07 Mbps, y un retardo de ping de 41 ms. Por otro lado, en el segundo test han aumentado las velocidades medias, en el enlace ascendente 15,19 Mbps, y en el enlace descendente 19,11 Mbps, y un retardo de 38 ms.

## VI. CONCLUSIONES

En el presente trabajo se ha conseguido implementar una estación base LTE utilizando el *software* libre Open Air Interface. Así, finalmente se han conseguido lograr los objetivos propuestos, realizando el estudio de distintos escenarios, así como la realización de distintas pruebas. En primer lugar, se ha implementado una red móvil celular LTE mediante dos máquinas virtuales, habiendo realizado la instalación y configuración en una de ellas de OASIM y éste asumiendo el papel de eNodeB + UE; mientras que en la otra se ha instalado y configurado el EPC y éste con las entidades de HSS + MME + S-GW.

En segundo lugar, fuimos un paso más allá queriendo trasladar la idea implementada en las MV a máquinas reales, utilizando un USRP y dispositivos móviles comerciales; permitiéndonos realizar un análisis exhaustivo de la parte radio, y también, se ha profundizado en el análisis de trazas.

En relación con el *software* utilizado hemos de comentar que OAI es una plataforma abierta y de código libre, y que se encuentra en desarrollo. Hemos encontrado algunas debilidades:

- La primera es que se trata de un paquete orientado a la investigación y la educación.
- En esta versión, OAI no soporta la conexión combinada, por lo que no podemos utilizar *Voice over LTE*.
- Inestabilidad al establecer unos RB superiores a 25 para así conseguir distintos anchos de banda (15 y 20 MHz), el MME termina su funcionamiento cuando el UE intenta conectarse a la red. Aunque pensamos que esta debilidad se puede deber a la falta de recursos por parte del equipo.

Para finalizar, este proyecto intenta plasmar bastante bien un escenario comercial, orientado a la capacitación laboral en el ámbito de las redes móviles de cuarta generación. En plena revolución de la tecnología 4G, y a las puertas de la 5G, la formación sobre redes de última generación promete un futuro alentador. Por otra parte, la finalización de este proyecto ha supuesto la satisfacción tanto a nivel personal como profesional, debido al total desconocimiento del funcionamiento y complejidad de una red móvil LTE.

### AGRADECIMIENTOS

Me gustaría aprovechar esta oportunidad, para agradecerle a mis padres Francisco y Carmen, por la educación que me han transmitido a mí y a mis hermanos. Ellos han conseguido de mí el hombre que soy. También, quiero agradecerle este mérito a mis hermanos Rafa y Eduvigis, siempre estaré para lo que necesitéis, a mis abuelos, a mis tíos, a mis primos, a mi familia, y a mi compañera de viaje Ana; ya que con su cariño y apoyo han logrado que hoy esté escribiendo estas palabras.

Además, me gustaría agradecerles el enorme apoyo que he recibido de mis tutores Jorge y José Carlos, por darme la oportunidad de realizar este proyecto con cara al futuro profesional. Siempre dispuestos a preocuparse por mis dudas; agradecerles sus consejos, su enseñanza, y por darme esa confianza.

### REFERENCIAS

- [1] 3GPP TS 36.104 v.10.2.0 Release 10, LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception.
- [2] openLTE An open source 3GPP LTE implementation, 2011. <https://sourceforge.net/p/openlte/wiki/Home> (Último acceso: 27/06/2017).
- [3] Página oficial de open air interface, 2017.
- [4] SDR Based LTE Workshop - United States Tour, 2017. <https://www.nutaq.com/lte-workshop-usa> (Último acceso: 20/07/2017).
- [5] CARDONA, N., OLMOS, J. J., GARCÍA, M., AND MONSERRAT, J. F. 3GPP LTE: hacia la 4G móvil, first ed. Marcombo Universitaria, 2011.
- [6] GUERRA, I. Conceptos básicos de telecos: del GSM al LTE, 2010. <https://blog.cnmec.es/2010/05/21/conceptos-basicos-del-telecos-evolucion-de-las-comunicaciones-moviles-del-gsm-al-lte> (Último acceso: 03/06/2017).
- [7] SHETTY, M. LTE as an all IP architecture, endorsed by many leading vendors, 2011. <http://www.kmshetty.com/2011/01/lte-as-all-ip-architectureendorsed-by.html> (Último acceso: 22/07/2017).
- [8] YACCHIREMA, C. 4G LTE definición, arquitectura, características, cálculos, simulaciones y seguridad, 2014. <http://inalambricas-lte4g.blogspot.com.es/2014/08/arquitectura-lte-la-arquitectura-lte.html> (Último acceso: 22/07/2017).
- [9] ZHENG, Q., DU, H., LI, J., ZHANG, W., AND LI, Q. Open LTE: an open LTE simulator for mobile video streaming. Paper presented at the I-2 (2014), 1–2.

# Análisis del registro espacial en imágenes de tomografía cerebral

Autor: Fernando Bayo Martínez, e-mail: nando.bm7@gmail.com

Tutor: Fermín Segovia Román, e-mail: fsegovia@ugr.es

Tutor: Diego Salas Gonzáles, e-mail: dsalas@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—El análisis de datos de neuroimagen entre diferentes pacientes requiere de un procesamiento previo, el cual asegura que el mismo vóxel en distintas imágenes se refiere a la misma posición anatómica. Varios algoritmos han sido propuestos para este fin, realizando la mayoría de ellos el registro espacial en dos pasos, una transformación afín seguida de un registro no lineal que posibilita la deformación de imágenes para ajustar en tamaño y forma las regiones individuales. Dicho registro no lineal puede provocar ruido que afecte negativamente a las imágenes neuronales.

En este proyecto se evalúa la influencia de la aplicación de transformaciones no lineales durante el registro espacial de neuroimágenes moleculares que serán utilizadas en sistemas informáticos destinados a ayudar al diagnóstico de trastornos neurodegenerativos. Concretamente, se compara el desempeño de un clasificador de máquina de vectores de soporte que hace uso de datos registrados espacialmente usando solo transformaciones afines y otro que emplea datos registrados usando el procedimiento clásico, incluyendo transformaciones no lineales.

**Palabras clave**—Alzheimer, imágenes de tomografía cerebral, máquina de vectores de soporte (SVM), Parkinson, regiones de interés (ROIs), registro espacial, validación cruzada.

## I. INTRODUCCIÓN Y MOTIVACIÓN

**D**URANTE las últimas décadas, el análisis de datos de imágenes neuronales se ha convertido en una práctica estándar para la evaluación inicial y la monitorización de pacientes que padecen trastornos neurodegenerativos. Un primer método consistía en el análisis visual de dichas imágenes por parte de médicos experimentados en busca de patrones específicos que caracterizaran las enfermedades bajo estudio. Más recientemente, se han propuesto una serie de sistemas de diagnóstico asistido por computadora (CAD) para analizar automáticamente estos datos y buscar patrones característicos, ayudando así al diagnóstico y seguimiento de enfermedades neurodegenerativas.

El uso de ordenadores en beneficio del análisis de datos de imágenes neuronales requiere algunos pasos previos adicionales destinados a hacer comparables los datos de distintos sujetos. Por ejemplo, los datos moleculares adquiridos mediante técnicas como la Tomografía Computarizada por Emisión de Fotón Único (SPECT) o de la Tomografía por Emisión de Positrones (PET), requieren un procedimiento de registro espacial (también conocido como normalización espacial) que modifica cada neuroimagen a fin de ajustarlas a una plantilla común previamente definida.

De los algoritmos propuestos para llevar a cabo el registro espacial [1], [2], [3], se observa como, en la mayoría de ellos,

se implementa dicho registro a través de dos procedimientos independientes: una transformación afín y un registro no lineal. El primero de ellos consiste en simples modificaciones tales como rotaciones y traslaciones, dejando la disposición cerebral dada sin cambios, mientras que el registro no lineal realiza transformaciones que modifican el tamaño y la forma de las regiones cerebrales individuales [4]. Aunque está ampliamente aceptado en la comunidad científica la necesidad de las transformaciones no lineales en el análisis de imágenes neuronales, estas producen fuertes deformaciones que pueden alterar considerablemente las imágenes originales. Recientemente, algunos estudios sugieren que estas modificaciones pueden distorsionar la información contenida en las imágenes y afectar negativamente al posterior análisis de la imagen, dando lugar a un diagnóstico erróneo de la enfermedad en estudio [5].

Es por este motivo que el objetivo principal de este trabajo reside en el análisis de la conveniencia de aplicar deformaciones no lineales en el registro espacial de datos de imágenes SPECT, comprobando si el ruido generado contrarresta el efecto positivo que la determinación de regiones específicas puede producir, a fin de mejorar el diagnóstico de enfermedades neurodegenerativas. Para ello, se utilizarán enfoques basados en el aprendizaje automático y se hará uso de dos conjuntos de datos, uno para el diagnóstico de la enfermedad de Alzheimer (EA) y otro para el diagnóstico de la enfermedad de Parkinson (EP).

Parte de los resultados de este proyecto han sido presentados en un congreso internacional organizado por el IEEE y publicados en su libro de actas [9].

## II. PLANIFICACIÓN

El esquema de la figura (1) muestra de manera resumida los pasos seguidos durante el desarrollo del proyecto. En este, se tratan las dos patologías neurodegenerativas más comunes en la actualidad: Alzheimer y Parkinson. Para poder trabajar con las bases de datos de imágenes correspondientes a pacientes y controles tanto de Alzheimer como de Parkinson, se debe realizar un primer paso de lectura y visualización de estas neuroimágenes, el cuál engloba el estudio de su formato (’.nii’), de los cortes (axial, coronal, sagital) y del software a utilizar.

Una vez hecho esto, comienza la etapa de pre-procesamiento, fase fundamental del proyecto pues el objetivo

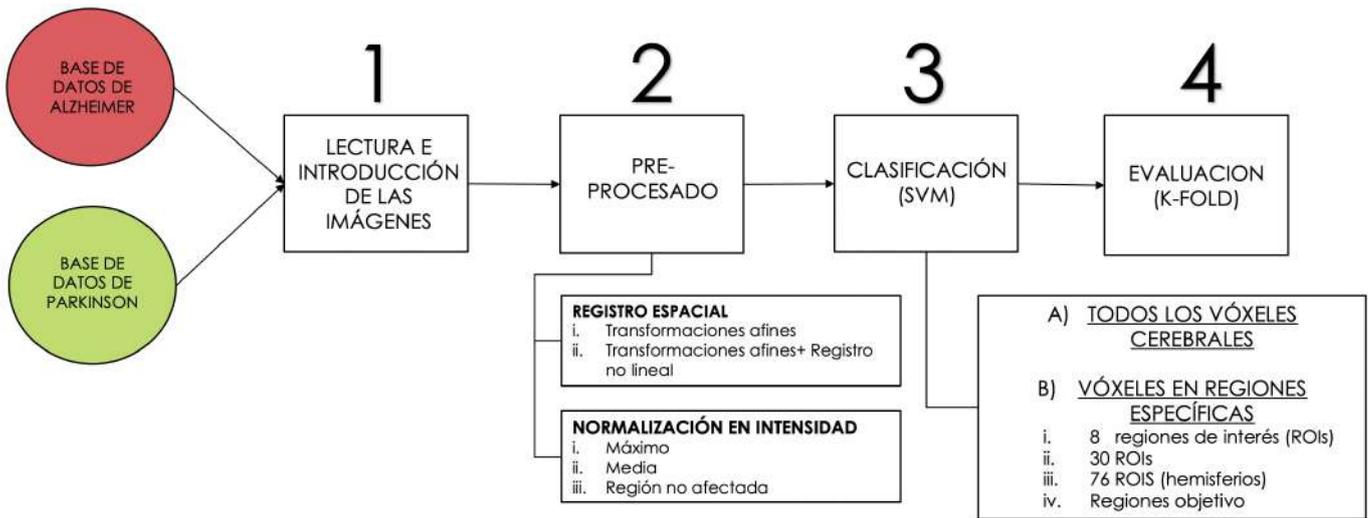


Fig. 1. Esquema del proceso llevado a cabo para el desarrollo del proyecto. Previa lectura e introducción de las bases de datos, se lleva a cabo un pre-procesamiento con el que hacer las imágenes comparables. En lo sucesivo, se realizan múltiples tipos de clasificación, evaluados mediante una técnica de uso extendido en inteligencia artificial conocida por *Validación cruzada K-fold*.

principal reside en el estudio del registro espacial. En esta etapa, se procede a la adecuación de las imágenes a fin de poder realizar una correcta clasificación posterior. Esta adecuación incluye el registro espacial o la normalización en intensidad, además de otros procedimientos tales como la corrección de la rotación apreciable a simple vista presente en las imágenes. En el caso del registro espacial, se analizará el efecto producido por la utilización de deformaciones no lineales y, para la normalización en intensidad, se estudiará qué método de los tres más usados (al máximo, a la media, a una región no afectada por la enfermedad) es el idóneo en cada caso, eliminando las variaciones de intensidad que no son debidas a diferencias en el estado patológico del sujeto. Finalizado el pre-procesamiento de las imágenes contenidas en las bases de datos, se realizan múltiples tipos de clasificación basados en algoritmos de aprendizaje automático que predicen la clase de una nueva muestra en función de un conjunto de datos entrenamiento. Así, se empleará un clasificador SVM para separar controles y pacientes utilizando dos criterios: (a) utilizando todos los vóxeles cerebrales como característica, situación en la que se considera el cerebro como una única región de interés y (b) únicamente utilizando los vóxeles de regiones específicas como característica. En este último caso, se estudiarán diferentes configuraciones como lo son la separación en 8, 30 o 76 (diferenciando regiones según se encuentran en el hemisferio cerebral derecho o izquierdo) regiones de interés (ROIs) o la clasificación a través de las regiones más afectadas por la enfermedad en cada caso que, a su vez, son las más discriminativas.

Por último, el modelo generado se evalúa mediante una técnica de uso extendido en inteligencia artificial conocida por *K-fold Cross Validation*. Con ella, se analiza el comportamiento del clasificador asegurando la independencia entre los datos del conjunto de entrenamiento y de test, a partir de los parámetros de precisión, sensibilidad y especificidad.

### III. MATERIALES Y MÉTODOS

#### A. Bases de datos

Las bases de datos empleadas han sido las siguientes:

- **Base de datos de EA:** 94 imágenes neuronales  $^{99m}\text{Tc}$ -ECD-SPECT pertenecientes a pacientes sanos y enfermos de EA adquiridas a partir de un estudio llevado a cabo por el hospital 'Virgen de las Nieves' (Granada, España). Tras el análisis, un grupo de experimentados médicos etiquetaron visualmente las imágenes como 'normal' (39) o 'EA' (55).
- **Base de datos de EP:** 200 imágenes neuronales  $^{123}\text{I}$  FP-CIT SPECT (DaTSCAN) pertenecientes a pacientes sanos y enfermos de EP obtenidas de la Iniciativa de Marcadores de Progresión del Parkinson (PPMI). Las imágenes han sido etiquetadas como 'normal' (99) o 'EP'(101).

#### B. Transformaciones afines

Este tipo de transformaciones permite identificar, para cada punto de una imagen  $[x_1, x_2, x_3]$ , un mapeo afín en las coordenadas de otro espacio  $[x'_1, x'_2, x'_3]$ , que se puede expresar en notación matricial como  $x' = \mathbf{M} \cdot x$ :

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \\ 1 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \quad (1)$$

A partir de la matriz  $\mathbf{M}$  se lleva a cabo la aplicación de traslaciones, rotaciones, zooms y cortes a fin de ajustar las neuroimágenes a una plantilla común. Para calcularla, se hace uso de un proceso iterativo el cual emplea un modelo afín de 12 parámetros y una función de coste dada por la diferencia cuadrática media entre la imagen original y la plantilla (2), la cual alcanzará su valor óptimo cuando ambas estén situadas en la misma posición anatómica.

$$F.O. = \sum_i (f(Mx_i) - g(x_i))^2 \quad (2)$$

Lo siguiente será utilizar el método de Gauss-Newton para estimar los parámetros  $q$  óptimos que minimizan la diferencia cuadrática anteriormente mencionada. Este método asume una función  $b_i(q)$  para definir la diferencia entre imágenes en un voxel específico  $i$ , usando parámetros de transformación  $q$ . Tras ello, mediante la primera aproximación del teorema de Taylor, se estima la diferencia entre imágenes para cada voxel cuando  $q$  se reduce en un factor  $t$  (3). Con base en esta ecuación, los valores de  $t$  que minimizan el error cuadrático medio dado por  $\sum_i b_i(q-t)^2$  pueden estimarse [6].

$$b_i(q-t) \simeq b_i(q) - t_1 \cdot \frac{\partial b_i(q)}{\partial q_1} - t_2 \cdot \frac{\partial b_i(q)}{\partial q_2} - \dots \quad (3)$$

### C. Registro no lineal

La utilización de deformaciones no lineales se utiliza especialmente en el ajuste de imágenes pertenecientes a distintos sujetos, aprovechando la consideración de las diferencias cerebrales de tamaño y forma existentes entre ellos. Su implementación viene precedida de la transformación afín de 12 parámetros y, de forma similar a la anterior, este método también emplea una función objetivo para medir la bondad del registro, procediendo de forma iterativa en la búsqueda de la deformación que proporcione el mayor valor de la función objetivo.

Las transformaciones no lineales vienen dadas por la expresión (4). Se utiliza una combinación de funciones básicas de baja frecuencia para determinar las deformaciones. Las funciones base son separables y cada una es generada por el producto de tres funciones básicas unidimensionales (5).

$$\begin{aligned} x'_1(\mathbf{x}, \alpha) &= x_1 + u_1 = x_1 + \sum_{m=1}^M \alpha_{m1} \phi_m(\mathbf{x}) \\ x'_2(\mathbf{x}, \alpha) &= x_2 + u_2 = x_2 + \sum_{m=1}^M \alpha_{m2} \phi_m(\mathbf{x}) \\ x'_3(\mathbf{x}, \alpha) &= x_3 + u_3 = x_3 + \sum_{m=1}^M \alpha_{m3} \phi_m(\mathbf{x}) \end{aligned} \quad (4)$$

$$\phi_m(\mathbf{x}) = \phi_{m_3}(x_3) \cdot \phi_{m_2}(x_2) \cdot \phi_{m_1}(x_1) \quad (5)$$

Las componentes de baja frecuencia de la transformada discreta del coseno (DCT) tridimensional se emplean como funciones básicas (6)(7).

$$\phi_1(i) = \frac{1}{\sqrt{I}} (i = 1 \dots I) \quad (6)$$

$$\phi_m(i) = \frac{2}{\sqrt{I}} \cdot \cos\left(\frac{\pi(2i-1)(m-1)}{2I}\right) \quad \begin{matrix} i = 1 \dots I, \\ m = 2 \dots M \end{matrix} \quad (7)$$

### D. Máquina de Vectores de Soporte (SVM)

Para proceder con la clasificación, cada imagen quedará representada por un vector de características, dividiendo el total de estos vectores en dos grupos independientes, uno para entrenamiento y otro para test. A partir del conjunto de entrenamiento se realizará una estimación de los parámetros

del clasificador. Este clasificador será una función construida a partir de los datos de entrenamiento y sus respectivas etiquetas, con el objetivo de averiguar la etiqueta correspondiente a los datos de test. Se trata de un método de clasificación binaria que estima la clase de una nueva observación desconocida por medio de un hiperplano, trazado para tener la mayor distancia a las muestras de entrenamiento más cercanas de cualquier clase. Matemáticamente, el hiperplano se puede expresar como:

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x}_i + w_0 \quad (8)$$

donde  $\mathbf{w}$  es el vector de pesos y  $w_0$  es el umbral. Las nuevas observaciones se clasifican según el lado del hiperplano en el que se encuentran situadas.

## IV. IMPLEMENTACIÓN Y RESULTADOS

Para evaluar la conveniencia de incluir transformaciones no lineales durante el registro espacial, los dos conjuntos de datos descritos anteriormente son pre-procesados incluyendo y sin incluir el paso de registro no lineal. Además, se realiza un proceso de normalización de intensidad que permite comparar las imágenes con precisión (ver Fig. 2). Tras ello, se hace uso de un clasificador SVM para separar los controles y pacientes usando dos criterios: a) emplear todos los voxéles cerebrales como característica y, b) emplear voxéles en regiones objetivo específicas como característica. Las regiones objetivo se definieron como temporal medio, temporal lateral, precuneus, cingulado posterior, cingulado anterior, frontal, occipital, estriado, tálamo y cerebelo para EA [7] y estriado para EP [8].

Para realizar la agrupación de voxéles de las neuroimágenes (una vez pre-procesadas) en las diferentes regiones de interés, se ha hecho uso de un atlas (*Automated Anatomical Labeling*). En todos los casos, el rendimiento de la clasificación SVM se evalúa utilizando un esquema de validación cruzada  $k$ -Fold con  $k=10$ . Es conveniente fijar la semilla del generador de números aleatorios para poder replicar los resultados, pues hay que tener en cuenta el carácter aleatorio de la distribución de imágenes en 'entrenamiento' y 'test'. Para ello, se repetirá cada clasificación un total de 10 veces, cada una con una semilla diferente. De este modo, se obtendrán 10 valores de precisión (ACC), especificidad (SPE) y sensibilidad (SEN) para cada clasificación con los que calcular el error asociado dado por la desviación típica. Así, se consigue obtener el margen de error asociado a cada medida.

Los resultados se muestran en las tablas I (cuando se usaron todos los voxéles) y II (cuando solo se usaron voxéles en las regiones objetivo). Véase como las deformaciones no lineales aumentan la precisión en aproximadamente un 2%, excepto en el caso de las regiones objetivo para el conjunto de 'Datos de EP'. Las figuras 3, 4 y 5 muestran la precisión de cada región cerebral incluida en el análisis. Para obtener dichos resultados basta con entrenar el sistema utilizando exclusivamente los voxéles de la región determinada.

## V. CONCLUSIÓN Y LINEAS FUTURAS

La comparación de imágenes cerebrales pertenecientes a distintos pacientes e incluso propias del mismo paciente tomadas en instantes diferentes es un área de acción compleja

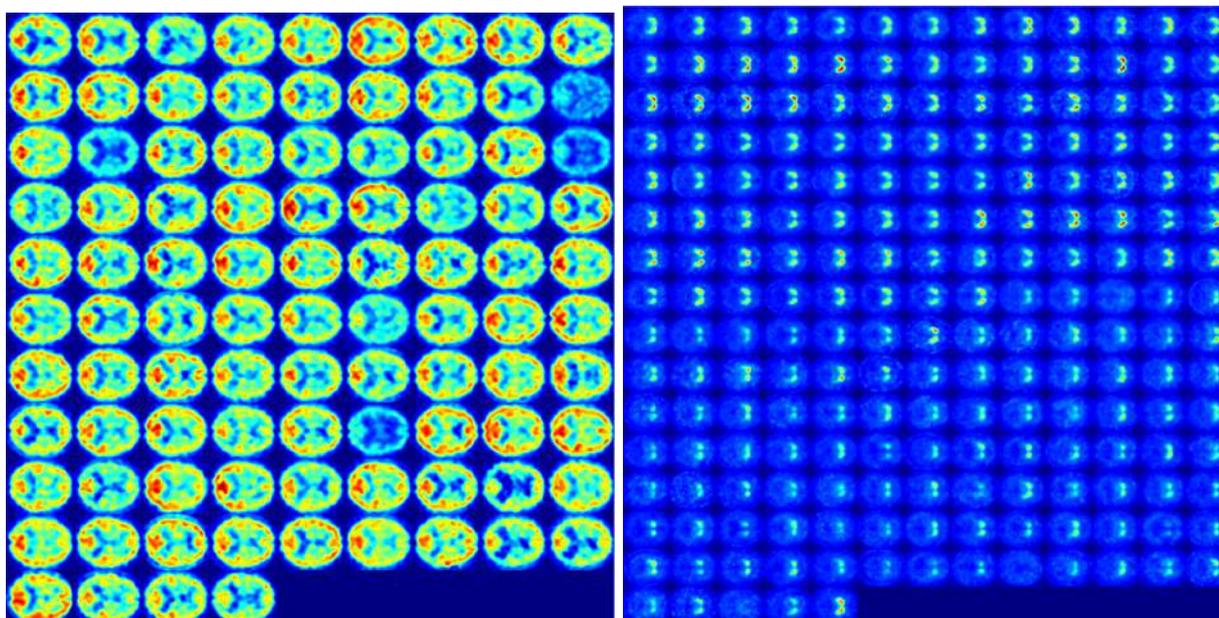


Fig. 2. Montaje del conjunto de imágenes de Alzheimer (izquierda) y Parkinson (derecha) tras la etapa de pre-procesamiento. Tras esta fase, la cual engloba el registro espacial y la normalización en intensidad, las imágenes son aptas para su comparación. Se puede observar como algunas de las imágenes están claramente afectadas por la enfermedad.

Tabla I  
RENDIMIENTO DE LA CLASIFICACIÓN EMPLEANDO TODOS LOS VÓXELES CEREBRALES

		ACC	SEN	SPE
<b>Base de Datos de EA</b>	Transf. Afines	83.29% ± 1.74%	85.45% ± 2.27%	80.25% ± 2.97%
	Afines + no lineales	85.32% ± 2.35%	86.00% ± 2.43 %	84.36% ± 4.10%
<b>Base de Datos de EP</b>	Transf. afines	95.80% ± 0.42%	97.88% ± 0.32%	93.76% ± 0.82%
	Afines + no lineales	97.40% ± 0.21%	97.98% ± 0.00%	96.83% ± 0.42 %

Tabla II  
RENDIMIENTO DE LA CLASIFICACIÓN CUANDO SE UTILIZAN COMO REGIONES OBJETIVO DE INTERÉS LOS LÓBULOS FRONTAL, PARIETAL, OCCIPITAL Y TEMPORAL PARA LA EA Y EL CUERPO ESTRIADO PARA LA EP.

		ACC	SEN	SPE
<b>Base de Datos de EA</b>	Transf. Afines	84.04% ± 1.66%	83.08% ± 4.05%	84.73% ± 1.27%
	Afines + no lineales	87.02% ± 1.79%	88.97% ± 2.43%	86.64% ± 2.34%
<b>Base de datos de EP</b>	Transf. Afines	97.45% ± 0.64%	97.48% ± 0.86%	97.43% ± 0.69%
	Afines + no lineales	97.20% ± 0.79%	97.07% ± 1.30%	97.33% ± 0.67%

a la par que necesaria para el diagnóstico e investigación de enfermedades actuales. Para llevar a cabo un análisis comparativo entre imágenes es necesario realizar un ajuste de estas en función de una plantilla común, proceso conocido por registro espacial. El principal interés perseguido en este proyecto ha sido la evaluación de la conveniencia de las transformaciones no lineales en dicho registro espacial. Para ello, se realiza un estudio de imágenes funcionales, pre-procesadas con y sin deformaciones no lineales, agrupadas en dos bases de datos. En vista de las dimensiones de las bases de datos, la relativa al Parkinson presenta un mejor ajuste de parámetros de la función de clasificación debido a su mayor tamaño, lo que hace que la precisión sea más alta que en el caso de la base de datos de Alzheimer. Del mismo modo, las imágenes DatSCAN obtendrán mejores resultados pues no

hay pacientes dudosos si no que tan solo se distingue entre sanos o enfermos.

Una vez finalizada la primera etapa de registro espacial se requiere de una normalización en intensidad, pues el nivel medio global del metabolismo cerebral también presenta variaciones entre diferentes imágenes. La elección de uno de los tres métodos posibles para este tipo de normalización (máximo, media y región no afectada) afecta, aunque no en gran medida, a los resultados finales. Por este motivo, se han ejecutado las tres técnicas sobre cada base de datos observando que los mejores resultados vienen proporcionados por una normalización al máximo en el caso del Alzheimer y una normalización a la media en el caso del Parkinson.

Los resultados obtenidos de la evaluación de la clasificación sugieren diversas interpretaciones según el tipo de

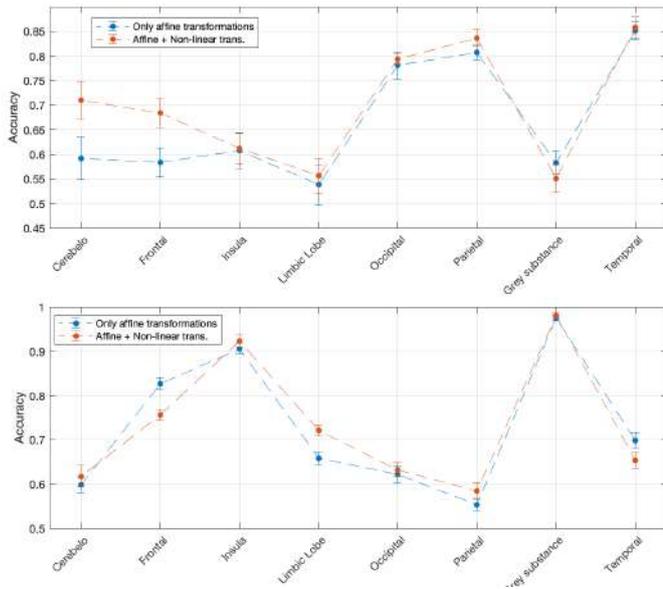


Fig. 3. Precisión alcanzada utilizando vóxeles agrupados en 8 ROIs como característica, al separar pacientes y controles con EA (gráfica superior), y pacientes y controles con EP (gráfica inferior).

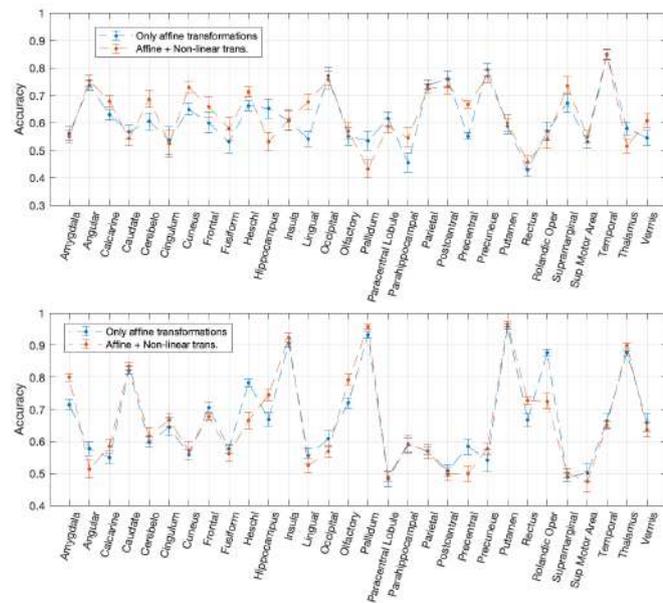


Fig. 4. Precisión alcanzada utilizando vóxeles agrupados en 30 regiones de interés como característica, al separar pacientes y controles con EA (gráfica superior) y pacientes y controles con EP (gráfica inferior).

clasificación implementada. Empleando todos los vóxeles del cerebro como características el resultado es claro, las transformaciones no lineales mejoran aproximadamente un 2% la precisión tanto en el caso del Alzheimer como en el caso del Parkinson. En el caso de clasificar utilizando únicamente los vóxeles de regiones específicas como características, los resultados son más ambiguos, siendo la tónica general la mejoría en la mayor parte de regiones de interés, a pesar de que algunas de las regiones más discriminativas se ven mejoradas prescindiendo del registro no lineal. Véase esto según las diferentes regiones tomadas:

- Tomando **8 ROIs** y clasificando independientemente

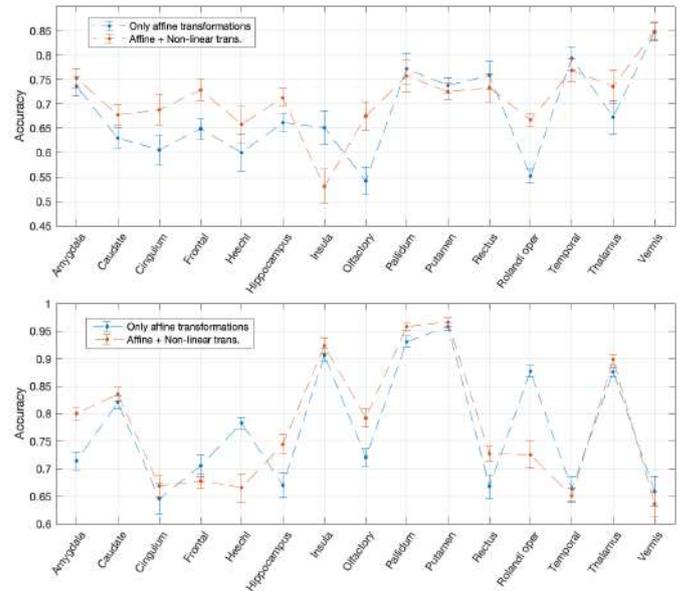


Fig. 5. Precisión alcanzada usando vóxeles en 76 regiones de interés como característica, al separar pacientes y controles con EA (gráfica superior) y pacientes y controles con EP (gráfica inferior). Debido a la extensión de los resultados obtenidos, se toman los más significativos, es decir, solo se seleccionan los resultados de precisión superior a 0,65%.

cada una de ellas, para ambos tipos de imagen se manifiesta una mejoría en los resultados tras la aplicación de deformaciones no lineales, si no en todas, en la gran mayoría de las clasificaciones (principalmente en las respectivas a las regiones más discriminativas).

- Tomando **30 ROIs** y clasificando independientemente cada una de ellas, al igual que en el caso anterior, se ha observado como la mayoría de las regiones (10 de las 15 regiones con precisión  $\geq 65\%$ ) toman las deformaciones no lineales como algo positivo en ambos tipos de imagen. Ahora bien, habría que tener en cuenta que, mientras que las regiones más discriminativas en el caso del Parkinson (estriado) se encuentran en este grupo, las que suelen verse más dañadas por la enfermedad de Alzheimer no mejoran su comportamiento a través del uso del registro no lineal.
- Diferenciando entre hemisferio izquierdo y derecho en las 30 ROIs anteriores se ha obtenido una nueva división en **76 ROIs**. En ese caso, tomando las regiones con precisión  $\geq 65\%$  aplicando algún tipo de registro espacial, se ha apreciado una mejoría dada por el registro no lineal. A pesar de que se trata de una diferencia existente en la mayoría de regiones, es muy leve en el caso de las imágenes DatSCAN y se ve acentuada sobre todo en las regiones con mayor acumulación de intensidad del estriado.
- Por último, se ha estudiado el resultado de la clasificación sobre 4 o 3 **regiones objetivo** para Alzheimer y Parkinson respectivamente. Finalmente se ha comprobado como esta clasificación experimenta una mejora en la precisión del 3% en el caso de imágenes ECD-SPECT, en contraposición al caso de imágenes DatSCAN, donde el resultado tras la aplicación de transformaciones no se ve alterado.

Consecuentemente, se puede afirmar que el uso de transformaciones no lineales mejora el registro de datos, haciéndolos más discriminativos, lo que permite la separación automática de pacientes y controles de forma más precisa.

Como trabajo futuro a corto plazo planeamos calcular si las diferencias en las tasas de precisión son estadísticamente significativas, asignando márgenes de confianza a las tasas. A mediano plazo, planeamos estudiar la influencia del registro no lineal en los procedimientos de extracción de características más comunes utilizados en los estudios de imagen neuronal.

#### AGRADECIMIENTOS

A Fermín Segovia, tutor de este Proyecto de Fin de Grado, por dedicar tu tiempo a orientarme y proporcionarme la ayuda y los medios necesarios durante este desafío. A todas aquellas personas que forman parte de mi día a día y al Departamento de Teoría de la Señal, Telemática y Comunicaciones de la ET-SIIT de la Universidad de Granada, por darme la oportunidad de dar a conocer este trabajo.

#### REFERENCIAS

- [1] K. J. Friston et al., *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, edición: 1 ed. Amsterdam ; Boston: Academic Press, Nov. 2006.
- [2] S. M. Smith et al., "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23 Suppl 1, pp. S208–219, 2004.
- [3] T. Puri et al., "A method for accurate spatial registration of PET images and histopathology slices," *EJNMMI Research*, vol. 5, Nov. 2015.
- [4] A. Klein et al., "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *NeuroImage*, vol. 46, no. 3, pp. 786–802, Jul. 2009.
- [5] D. Castillo-Barnes et al., "Comparison Between Affine and Non-affine Transformations Applied to I123-FP-CIT SPECT Images Used for Parkinson's Disease Diagnosis," in *Understanding the Brain Function and Emotions*, ser. Lecture Notes in Computer Science, J. M. Ferrández Vicente et al., Eds. Almería, Spain: Springer International Publishing, May 2019, pp. 379–388.
- [6] J. Ashburner et al., "Chapter 04 - Rigid Body Registration," in *Statistical Parametric Mapping*, K. Friston et al., Eds. London: Academic Press, 2007, pp. 49–62.
- [7] F. Segovia et al., "Usefulness of Dual-Point Amyloid PET Scans in Appropriate Use Criteria: A Multicenter Study," *Journal of Alzheimer's Disease*, vol. 65, no. 3, pp. 765–779, Jul. 2018.
- [8] F. Segovia et al., "Preprocessing of 18F-DMFP-PET Data Based on Hidden Markov Random Fields and the Gaussian Distribution", *Frontiers in Aging Neuroscience*, vol. 9, pp. 1–9, Oct. 2017.
- [9] F. Bayo et al., "Spatial registration of neuroimaging data: Analysis of the convenience of performing non-affine transformations", *Virtual 2020 IEEE Nuclear Science Symposium & Medical Imaging Conference*, Boston (USA), 31 Oct. – 7 Nov. 2020.



**Fernando Bayo Martínez.** Nacido en Granada en 1998. Graduado en Ingeniería de Tecnologías de Telecomunicación por la Universidad de Granada en 2020.

# Diagnóstico de la enfermedad de Alzheimer basado en el análisis de imágenes médicas cerebrales mediante redes neuronales.

Autor: Pablo Manresa Nebot, e-mail: pablo.96.almo@gmail.com

Tutor: Fermín Segovia Román, e-mail: fsegovia@ugr.es

Titulación: Ingeniería Informática

Departamento de Teoría de la Señal y Comunicaciones

Universidad de Granada

**Resumen**—En este trabajo, se propone la creación de un sistema el cual utiliza dos enfoques, el primero, se basa en la clasificación de imágenes de los pacientes, produciendo como salida un diagnóstico. Se tienen 3 clases; Alzheimer, Normal y MCI. En cambio, el segundo, usa un enfoque basado en la regresión, y permite establecer una correlación entre el valor asociado del Mini-Mental State Examination (MMSE) y la tomografía del paciente, proporcionando así, una monitorización. En total se dispone de 26 etiquetas diferentes, que van del valor 4 al 30. Para entrenar los modelos se usan imágenes en formato MRI tanto 2D como en 3D. El paradigma está basado en Deep Learning ya que las arquitecturas usadas tanto para la regresión como para la clasificación son redes neuronales convolucionales. Para el enfoque de diagnóstico se consigue una precisión cercana al 91% frente a la monitorización, que consigue un error cuadrático de 6.8.

**Palabras clave**—Alzheimer, Clasificación, Deep Learning, Enfermedades Neurodegenerativas, Keras, Mmse, Nibabel, Opencv, Pandas, Redes Convolucionales, Reducción Espacial, Regresión, Resnet, Sklearn, VGG, Vóxeles.

## I. INTRODUCCIÓN

**E**L Alzheimer es una enfermedad de lento desarrollo, presentando leves problemas cognitivos y finalizando con un grave daño cerebral. La esperanza de vida de un paciente con Alzheimer es de 8 a 10 años. Durante la última década, durante el diagnóstico estaba presente el factor humano y la subjetividad inherente que ello conllevaba, ya que éste era realizado por parte de un especialista. Para mejorar el diagnóstico y solventar la problemática de la subjetividad, se propone un sistema basado en Deep Learning, que usando imágenes por resonancia Magnética (MRI) proporciona un diagnóstico o monitorización con alto grado de objetividad. De este modo, el sistema realiza el trabajo de un médico especialista pero solventando el problema anterior.

Este sistema no solo está pensado para asistir a los médicos, sino que también permite monitorizar la progresión de la enfermedad, y así asignar un número (del 1 al 30) al estado del paciente, siendo el 30 el que mejor estado presenta y 1 el que peor [1].

## II. METODOLOGÍAS PROPUESTAS

Las imágenes inicialmente se presentan en formato 3D, es decir, con las dimensiones (157, 189, 156), esto puede suponer un problema para computadores que no son de altas prestaciones debido a su alta dimensionalidad. En este trabajo se proponen varias heurísticas para reducirla, la primera consiste en reducir a la mitad la dimensionalidad de las imágenes en cada uno de los ejes, y usar la media como heurística para seleccionar el nuevo bloque. Esto permite reducir la dimensionalidad de (157, 189, 156) a (79,95, 78). Por otro lado, se presenta una heurística, la cual consiste en coger cortes (axial, coronal o sagital) 2D en cada una de las imágenes para un determinado eje. Para este trabajo se han cogido 20 cortes axiales por cada imagen 3D, en aquellas zonas donde la información discriminante era mayor. Las dimensiones pasan de (157, 189, 156) a (157, 189) por cada imagen en 2D. Sin embargo, la pérdida de información es mayor. En este trabajo se ha hecho un estudio con ambas heurísticas para cada uno de los enfoques para comprobar cual proporciona mejor resultado.

## III. DATASET

Las imágenes MRI han sido extraídas de la base de datos de Alzheimer's Disease Neuroimaging Initiative (ADNI [4]). Por un lado se presenta una versión con 909 imágenes, y por otro, una versión reducida de éste con 428. La versión de 909 se utilizará para los problemas que usen datos en 2D, mientras que la versión reducida será para los problemas en 3D.

### A. Procesado

En cuanto al procesado de las imágenes, éstas se segmentaron usando el algoritmo unificado, el cual usa la mezcla de Gaussiana extendiéndose para incorporar una variación suave de intensidad junto a un registro no lineal con mapas de texturas. Se descartaron todos los tejidos diferentes de la materia gris.

### B. Clasificación

Para el problema de clasificación se dispone de las clases, Alzheimer's disease (AD), Normal (NOR) y Mild Cognitive Impairment (MCI). La distribución de datos para la clasificación:

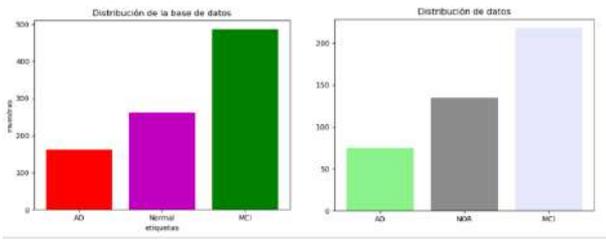


Fig. 1. Distribución de datos para la clasificación. Dataset completo vs. Dataset reducido.

C. Regresión

Para la regresión, en función de la versión elegida del dataset, se posee un número de clases u otro. Las diferentes clases corresponden a los numerosos valores del MMSE. La distribución de datos es la siguiente:

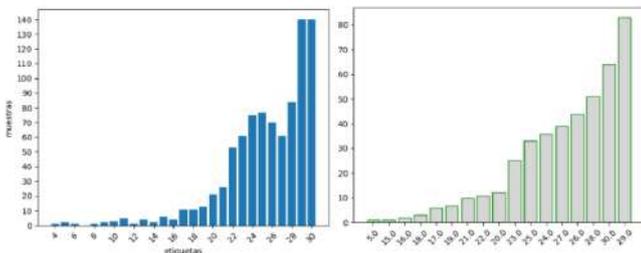


Fig. 2. Distribución de datos para la regresión. Dataset completo vs. Dataset reducido.

Como puede observarse, la distribución de datos tanto para la regresión como para la clasificación está bastante desbalanceada, lo cual evidencia la dificultad de realizar trabajos en este tipo de campos, ya que la mayor parte de las veces, se dispone de datasets incompletos, desbalanceados y sobre todo, con pequeñas cantidades de muestras. En el caso de la regresión todavía es más evidente puesto que hay clases como la 4, o la 6, dónde únicamente hay un par de fotos. Esto supone un gran problema, ya que las redes neuronales funcionan especialmente bien cuando la cantidad de datos es elevada.

IV. ARQUITECTURAS

A. Clasificación 2D

En primer lugar esta red presenta dos bloques convolucionales consecutivos. Cada uno de los bloques está compuesto por una capa convolucional, otra de tipo BatchNormalization, a continuación, otra de activación y finalmente, una de tipo MaxPooling. El esquema es el siguiente:



Fig. 3. Bloque de tipo convolucional.

Posteriormente, la arquitectura presenta dos bloques consecutivos de tipo Inception [2], es decir, bloques con varias capas en paralelo, que usan convoluciones de 1x1 de este modo, permiten reducir la dimensionalidad, y así poder crear redes más densas. El resultado de las capas anteriores se concatena y se pasa como entrada a la siguiente capa. El

esquema es el siguiente:

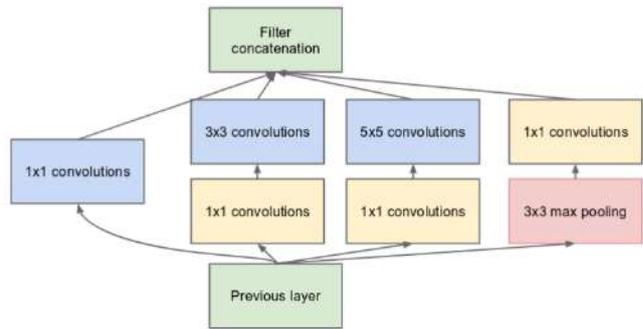


Fig. 4. Bloque de tipo Inception.

Tras la salida del segundo bloque de tipo Inception [2], el siguiente bloque, es de tipo Residual, el cual permite crear saltos entre bloques, y resolver problemas como el famoso desvanecimiento del gradiente. La versión usada en esta arquitectura consta de dos arcos, el primero que posee dos capas convolucionales consecutivas, y el segundo, que únicamente posee una. Finalmente, el resultado de ambas capas se suma. El esquema:

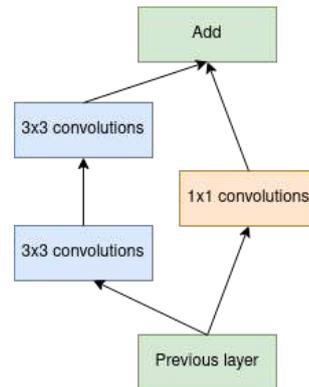


Fig. 5. Bloque de tipo Residual.

A continuación, se presenta una capa de activación seguida de un último bloque convolucional. El último bloque es de tipo denso, y está compuesto por una capa de tipo densa, una de tipo dropout, para así evitar el sobreajuste y finalmente, otra de tipo densa con 3 neuronas (3 clases) y como función de activación Softmax.

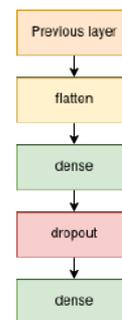


Fig. 6. Último bloque de la red.

### B. Regresión 3D

Esta arquitectura es bastante simple en comparación al resto, puesto que está compuesta de 3 bloques convolucionales, y uno de tipo denso. Diseccionando el bloque convolucional, se tiene, en primer lugar, una capa de tipo convolucional, en segundo se tiene una capa de tipo Batch-Normalization para normalizar los datos, posteriormente una capa de activación, y finalmente, una capa de MaxPooling. El esquema:

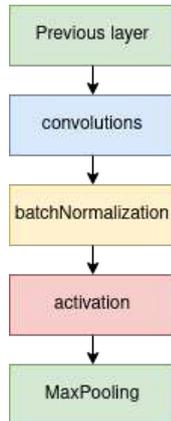


Fig. 7. Bloque convolucional.

En cuanto al bloque denso, su composición consta de una capa de tipo densa o fully connected, una capa de dropout para prevenir el sobreajuste y finalmente, otra capa densa con su respectiva función de activación. Al ser un problema de regresión, será de tipo lineal. El esquema es similar al de la Fig. 6.

### C. Otros enfoques

En cuanto a la clasificación 3D, la arquitectura usada está basada en el modelo ResNet-18 [3], es decir, está compuesta por 18 capas. En cambio, para el enfoque de regresión 2D, el modelo usado está basado en Resnet-50 [3]. El motivo por el que no se han descrito de forma detallada las arquitecturas es debido a que, como se expone más adelante, son los enfoques para cada problema que peor resultado han proporcionado.

## V. MÉTRICAS

En cuanto a las métricas utilizadas para evaluar la calidad los modelo, se dispone de varias en función del tipo de problema.

### A. Error cuadrático medio

Por un lado de la métrica del error cuadrático medio para el problema de monitorización (regresión), y calcula la media de las diferencias entre las clases predichas y las reales, y las eleva al cuadrado para impedir que sea un número negativo. Esto es:

$$mse = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

donde  $\hat{y}$  es la clase predicha e  $y$  la clase real.

### B. Exactitud

Esta métrica es utilizada para el problema de diagnóstico (clasificación), y permite calcular el número de muestras que han sido clasificadas correctamente con respecto al total. La fórmula sería:

$$exactitud = \frac{prediccionescorrectas}{predicciones totales} \times 100 \quad (2)$$

### C. Matriz de Confusión

Finalmente, la matriz de confusión es una métrica utilizada en el problema de clasificación, y permite conocer con un alto grado de detalle las predicciones que el modelo ha realizado, es decir, los falsos positivos y negativos, así como los verdaderos positivos y negativos. De esta forma, es posible actuar sobre ellas e intentar minimizar aquellas cuyas consecuencias pudieran ser trágicas. En un sistema de diagnóstico, interesa minimizar los falsos negativos todo lo posible, puesto que podrían empeorar o, incluso, acortar la vida del paciente. Su estructura sería:

		real		
		AD	MCI	NOR
Predicción	AD	117	20	13
	MCI	7	333	18
	NOR	1	20	199

Fig. 8. Matriz de confusión.

## VI. ENTRENAMIENTO Y AJUSTE DE HIPERPARÁMETROS

En cuanto al entrenamiento, se optó por uno basado en validación cruzada con  $k$  particiones, esto es, se divide el conjunto en  $k$  particiones, y se usa cada una de ellas para validar el modelo. Los datos contenidos en la partición de validación no han de estar en la de entrenamiento. En cuanto a los optimizadores tanto para la clasificación 2D como para la regresión 3D se ha usado el gradiente descendente estocástico. Para la clasificación, el ratio de aprendizaje ha sido de 0.001 y la función de pérdida ha sido categorical crossentropy. En el caso de la regresión 3D, la función de pérdida ha sido el error cuadrático medio y el ratio de aprendizaje de 0.0001.

Además, se han usado monitores para serializar los datos, y para que el modelo pare de entrenar en caso de que no haya una mejora en X iteraciones.

## VII. RESULTADOS

### A. Clasificación

En cuanto a la clasificación 2D, se optó por un modelo basado en validación cruzada con 5 particiones, de los cuales, se reservaron un 80% de los datos para la validación y el entrenamiento, mientras que un 20% fueron reservados para el test. Los resultados obtenidos para la clasificación:

Tabla I  
PRECISIÓN PARA CADA UNA DE LAS PARTICIONES.

Partición	exactitud
1	0.923
2	0.895
3	0.928
4	0.902
5	0.901
Total	0.91

En cuanto a la matriz de confusión:

Tabla II  
MATRIZ DE CONFUSIÓN PARA LA CLASIFICACIÓN 2D.

		AD	Real MCI	NOR
Predicha	AD	80.32%	1.67%	1.31%
	MCI	14.51%	95.78%	8.89%
	NOR	5.17 %	2.56%	89.80%

Finalmente, la comparativa entre el modelo de clasificación 2D y el modelo de clasificación 3D:

Tabla III  
MODELO 2D vs. MODELO 3D.

	exactitud
Modelo 2D	0.9098
Modelo 3D	0.731

### B. Regresión

Finalmente, para la regresión, la división de los datos fue de un 85% para train y validación y un 15% para test. Los resultados obtenidos:

Tabla IV  
MODELO 2D vs. MODELO 3D.

	MSE
Modelo 2D	7.712
Modelo 3D	6.829

## VIII. CONCLUSIÓN Y TRABAJO FUTURO

Se presenta un sistema que con un 91% de exactitud proporciona un diagnóstico a un paciente. Además, presenta un 6.8 de MSE durante la monitorización (recordemos que es cuadrático y el número de clases va del 4 al 30). Para el caso de la regresión, el problema es especialmente difícil puesto que se dispone de una gran cantidad de clases pero un número muy pequeño de muestras para la mayoría de ellas (la mayor cantidad de muestras se concentran en las clases 29 y 30). Para mitigar el anterior problema, se propone como continuación de este trabajo, la creación de una red neuronal generativa adversaria, que permita crear imágenes sintéticas cerebrales, y de este modo, poder ampliar el dataset especialmente con las clases infrarrepresentadas, tales como la clase AD, o la mayoría de las clases con valores numéricos diferentes a 29 o 30.

Parte de este trabajo ha sido presentado en el congreso internacional **IEEE Nuclear Science Symposium (NSS) and Medical Imaging Conference (MIC)** y publicado en su libro de actas con el **ID 1800** y el título **Estimating the severity of Alzheimer’s disease using Convolutional Neural Networks and Magnetic Resonance Imaging Data** [5].

### AGRADECIMIENTOS

En primer lugar, me gustaría agradecerlo a mis padres, por hacer posible que pudiera completar mis estudios y haber hecho de guía y apoyo en los momentos más difíciles. En segundo lugar, me gustaría agradecerlo a mis abuelos, ya que ellos me han inculcado una serie de valores que han hecho que me convirtiera en la persona que hoy en día soy.

### REFERENCES

- [1] Jane Upton. “Mini-Mental State Examination”. In: *Encyclopedia of Behavioral Medicine*. Ed. by Marc D. Gellman and J. Rick Turner. New York, NY: Springer New York, 2013, pp. 1248–1249. ISBN: 978-1-4419-1005-9. DOI: 10.1007/978-1-4419-1005-9\_473. URL: [https://doi.org/10.1007/978-1-4419-1005-9\\_473](https://doi.org/10.1007/978-1-4419-1005-9_473).
- [2] Christian Szegedy et al. “Going Deeper with Convolutions”. In: *CoRR* abs/1409.4842 (2014). arXiv: 1409.4842. URL: <http://arxiv.org/abs/1409.4842>.
- [3] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV].
- [4] the ADNI team. *ADNIMERGE: Alzheimer’s Disease Neuroimaging Initiative*. 2020.
- [5] P. Manresa-Nebot et al. “Estimating the severity of Alzheimer’s disease using Convolutional Neural Networks and Magnetic Resonance Imaging Data”. In: (2020).

# Diseño e implementación de modelos que combinan diferentes modalidades de imagen médica cerebral para mejorar el diagnóstico asistido de la enfermedad de Alzheimer

Autor: Sergio Ortiz Ruiz , e-mail: Sergioortiz1@correo.ugr.es

Tutor: Fermín Segovia Román , e-mail: fsegovia@ugr.es

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones

Universidad de Granada

**Resumen**—El presente proyecto se basa en analizar diferentes modalidades de imágenes cerebrales para mejorar el diagnóstico asistido de la enfermedad de Alzheimer (AD). Las imágenes son analizadas para discernir entre sujetos sanos (NOR), pacientes con enfermedad de Alzheimer (enfermos\_AD) o pacientes con deterioro cognitivo leve (MCI). Se utilizan dos modalidades de imagen cerebral: resonancia magnética (MRI) y tomografía por emisión de positrones (PET). El análisis se realiza en distintas etapas. En primer lugar se estudian las imágenes PET por un lado y las imágenes MRI por otro, y en segundo lugar se estudia la combinación de ambas modalidades de imagen. El proyecto analiza la influencia del uso combinado de estas imágenes mejora la detección de la enfermedad.

El análisis computarizado de las imágenes requiere de un preprocesado previo de las mismas que las haga comparables entre sí. Para ello se aplican métodos de normalización de las imágenes. Una vez que las imágenes han sido preprocesadas y almacenadas correctamente se realiza el proceso de clasificación. La clasificación se lleva a cabo mediante modelos de aprendizaje automático supervisado, principalmente mediante máquinas de vectores soporte (SVM). La evaluación del rendimiento de los clasificadores se realiza a través del método de validación cruzada k-fold, comprobando las estimaciones con las etiquetas originales de los datos.

**Index Terms**—Aprendizaje automático, máquinas de vectores soporte, validación cruzada, reducción de dimensionalidad, Alzheimer, neuroimagen, enfermedades neurodegenerativas, demencia.

## I. INTRODUCCIÓN

Las enfermedades neurodegenerativas, y principalmente la enfermedad del Alzheimer (AD), se han convertido en una de las patologías más notables en las últimas décadas, siendo el diagnóstico precoz de estas enfermedades de vital importancia. La AD es la demencia con más prevalencia del mundo y concentra entre el 60 y el 80 por ciento de los casos.[5]

El principal factor de la AD es la pérdida de memoria, aunque produce otras alteraciones en los enfermos como la pérdida de orientación, problemas con el lenguaje o alteraciones en el carácter. La enfermedad evoluciona en diversas etapas provocando daños cerebrales y causando la muerte. El diagnóstico de AD es complicado y subjetivo por naturaleza en una gran cantidad de casos, siendo difícil diferenciar en ocasiones cual es el límite que separa una

persona enferma o sana. En cuanto a los cambios que se producen a nivel cerebral en las demencias, el deterioro afecta de manera selectiva en diversas zonas.

En las últimas décadas la tecnología de imagen médica ha sufrido un gran desarrollo, siendo de gran ayuda para detectar enfermedades neurológicas como las demencias. Las técnicas de imagen médica son técnicas no invasivas, es decir, no producen alteraciones en los pacientes evitando un proceso doloroso. A través de análisis realizados a estas imágenes con sistemas informáticos, se intentan localizar patrones que proporcionen la información para detectar el estado de un paciente. Estos sistemas informáticos se basan, en muchos casos, en algoritmos de aprendizaje automático. Estos métodos son los denominados sistemas de diagnóstico asistido por ordenador (CAD), sirviendo estos como apoyo al diagnóstico médico.

El análisis computarizado de las imágenes requiere de un preprocesado previo de las mismas que las haga comparables entre sí. Una vez que las imágenes han sido preprocesadas y almacenadas, se encuentran adecuadas para la clasificación que se realiza en este proyecto mediante modelos de aprendizaje automático supervisado, principalmente a través de máquinas de vectores soporte (SVM). Debido al alto número de características que contiene cada representación cerebral, se estudia la reducción de dimensionalidad de las imágenes, como mejora de los rendimientos. La evaluación del rendimiento de los clasificadores se realiza a través del método de validación cruzada, comprobando las estimaciones con las etiquetas originales de los datos.

## II. MATERIALES Y MÉTODOS

### II-A. BASE DE DATOS

Las imágenes que se utilizan durante la realización del proyecto forman parte de la Iniciativa de neuroimagen de la enfermedad de Alzheimer (ADNI, Alzheimer's Disease Neuroimaging Initiative). El proyecto ADNI fue diseñado para desarrollar mecanismos de detección de la AD principalmente en etapas tempranas y almacenar datos referentes a la propia enfermedad. La base de datos de ADNI permite la disposición de información a investigadores de todo el

mundo, para desarrollar estudios y mejorar en conjunto los diagnósticos de la enfermedad.

La base de datos de nuestro proyecto esta formada por 249 sujetos, de los que tenemos imágenes cerebrales tanto estructurales MRI como funcionales PET. Las imágenes MRI proporcionan un examen imageneológico mediante el empleo de imanes y ondas de radio. Estos producen la creación de un campo magnético en el cual los protones del cuerpo de alinean con el campo. Las imágenes PET conforman la segunda técnica que tratamos en el proyecto, permitiendo obtener imágenes mediante el uso de medicina nuclear. Las imágenes son generadas por la captación de positrones, siendo estos emitidos en el proceso de descomposición de los isótopos radiactivos que incorpora el radiofármaco. Los 249 sujetos están organizados de la siguiente forma:

- 68 pertenecen a sujetos controles sanos (NOR).
- 111 pertenecen a sujetos con deterioro cognitivo leve (MCI).
- 70 pertenecen a sujetos con AD (enfermos\_AD).

## II-B. PREPROCESADO DE LAS IMAGENES

### Normalización espacial

Las estructuras cerebrales son organizaciones complejas, y debido a la anatomía de cada paciente hace difícil la comparación de imágenes sin estos procesos de corrección. La normalización espacial es la fase donde las imágenes son adaptadas en cuanto a su forma según una plantilla común, situando esto las imágenes de entrada de diferentes pacientes en un mismo espacio. La plantilla común se corresponde con una imagen cerebral que representa una anatomía estándar. Se utiliza en el proceso de normalización el algoritmo de información mutua. Los procesos que se llevan a cabo conllevan la aplicación de transformaciones tanto lineales como no lineales. La función de optimización presenta el valor óptimo cuando la posición anatómica de la imagen y la plantilla son iguales.

$$f_{obj} = \sum_i (f(x_i) - g(x_i))^2 \quad (1)$$

Algunos de los procesos que se pueden realizar durante esta fase son traslación, rotación y deformaciones de áreas cerebrales.

### Normalización de la intensidad

El siguiente proceso es lo que se conoce como normalización de intensidad. Este paso se realiza en imágenes PET donde la información más útil se encuentra en los niveles de intensidad de los vóxeles y el propósito fundamental es igualar la intensidad entre pacientes. Estas diferencias en cuanto a las intensidades de las imágenes pueden deberse a razones como el uso de diferentes dosis de fármaco, diferencias en los efectos del fármaco en el cuerpo de los pacientes o la utilización de difentes cámaras para la adquisición de imágenes entre pacientes. En el trabajo se implementa la normalización a valor máximo, consistiendo este proceso en la normalización de cada voxel de la imagen en función del valor medio 1% de los vóxeles de mayor

intensidad, dividiendo cada uno de ellos por el valor medio.

### Segmentación de las imágenes estructurales.

El proceso de segmentación en las imágenes estructurales separa la representación inicial en los diferentes tejidos que las componen, para realizar un análisis más exhaustivo de estos tejidos. Se realiza una división de tejido cerebral, estando formado por la materia gris, materia blanca y líquido cefaloraquídeo, del resto de tejidos de la cabeza. La materia gris principalmente es aquella formada por los conjuntos de cuerpos neuronales, mientras que la materia blanca corresponde a axones, o fibras nerviosas. La segmentación se logra mediante el uso de mapas de probabilidad de tejido, que cuantifican la probabilidad de la presencia de cierto tipo de tejido para cada vóxel.

## II-C. REDUCCIÓN DE LA DIMENSIONALIDAD

Una alta dimensionalidad de los datos no siempre proporciona los mejores resultados. En ocasiones, para una base de datos pequeña, la alta dimensionalidad de las imágenes no proporciona los mejores resultados. Por esta razón, este problema se resuelve utilizando métodos de reducción de dimensionalidad que se enfocan en la información útil de cada imagen. Durante este proyecto se utilizan dos técnicas:

**Selección de las características más importantes.** Este método se basa en usar las características clave por criterios de separabilidad de clases. El fundamento principal es eliminar aquellas variables menos importantes, seleccionando aquellas características de los objetos cuyos pesos son más altos a la hora de clasificar y que definen mejor la diferenciación de clases. Para discernir entre las características se utiliza el método de varianza agrupada, donde se estudian las características de las imágenes tomando aquellas cuya deferencia entre clases sea más alta.

**Principal Component Analysis (PCA).** PCA es un procedimiento matemático que posibilita rotar los ejes del espacio de los datos sobre las líneas de máxima varianza, permitiendo con ello minimizar la complejidad de espacios con alta dimensionalidad sin eliminar mucha información. Los ejes de mayor varianza se denominan componentes principales. Teniendo las imágenes almacenadas en una matriz  $X=[X_1, X_2, \dots, X_N]$  siendo  $N$  el número de sujetos, se normaliza el valor de esta matriz a la norma restando la media, obteniendo con ello la matriz  $Y=[Y_1, Y_2, \dots, Y_N]$ . El siguiente paso es calcular la matriz de covarianza de la matriz  $Y$  : [6]

$$C = cov(Y, Y) = \frac{1}{N} * Y Y^t \quad (2)$$

Calculada esta matriz de covarianza obtendremos los valores propios  $\Lambda$  y los vectores propios  $\Gamma$  asociados. Una vez que estos son calculados, los datos se reproducen sobre los vectores propios, sobre todos o sobre los que consideremos más importantes. Los vectores propios son unitarios y perpendiculares entre sí, y nos proporcionan información sobre los patrones en los datos.

## II-D. APRENDIZAJE AUTOMÁTICO

El aprendizaje automático es una materia de la inteligencia artificial donde buscamos construir una serie de programas que tengan capacidad de aprender, basándose directamente en la experiencia. El trabajo se basa en un aprendizaje automático supervisado, donde los datos que usamos para entrenar nuestro clasificador deben estar etiquetados. El proceso de clasificación se basa en la utilización de dos conjuntos de entrada, denominados conjunto de entrenamiento (train) y conjuntos de test (test). Estos conjuntos de entrenamiento y test son obtenidos a través del método de validación cruzada K-fold.

- K-fold validación cruzada.** Procedimiento que divide el conjunto inicial  $M$  de datos en  $k$  subconjuntos equitativos, de los cuales en el proceso de entrenamiento del clasificador se tomarán  $k-1$  conjuntos y en el proceso de evaluación se tomará el conjunto restante. El proceso de elección de los conjuntos se repetirá  $k$  veces, cogiendo en cada iteración un conjunto diferente para test del clasificador y los  $k-1$  conjuntos restantes para el entrenamiento.

Se van a estudiar varios casos de clasificadores, como son los clasificadores  $K$  vecinos mas cercanos (KNN), árboles de decisión (TREE) o máquinas de vectores soporte (SVM) para imágenes cerebrales, centrándonos principalmente en el método SVM.

**SVM.** Este método de clasificación se basa en la creación de un hiperplano de separación entre datos. Este hiperplano realiza la función de divisor, separando el espacio  $n$ -dimensional en dos mitades donde cada clase de objetos queda a un lado.

Un hiperplano puede ser descrito por una ecuación lineal de la forma:

$$w_1x_1 + w_2x_2 + \dots + w_nx_n = b \quad (3)$$

Las SVM quedarán definidas principalmente por 3 factores [1]:

- Hiperplano  $\rightarrow$  superficie de separación
- Vectores de soporte  $\rightarrow$  puntos mas cercanos al hiperplano principal que definen los hiperplanos paralelos.
- Margen máximo  $\rightarrow$  banda o rango de valores que proporcionan la separación entre los datos de una de las clases y la otra

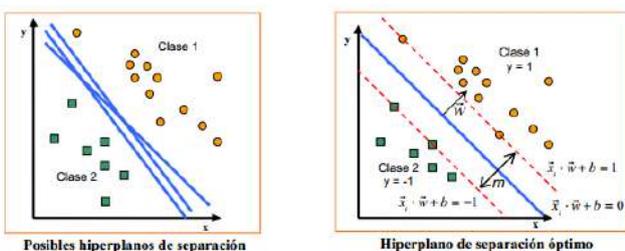


Figura 1. Hiperplanos de separación SVM.

Por tanto para conseguir que el margen de decisión sea lo más ancho posible la cuestión pasa por maximizar  $\frac{2}{\|w\|}$  para obtener el vector  $w$  que define al hiperplano [7]. Los puntos que se encuentran mas adyacentes a la frontera de decisión (vectores soporte), serán aquellos que dan lugar a dos hiperplanos paralelos al hiperplano principal ( $h_1$  y  $h_2$ ).

Debido a que la función del hiperplano tiene  $n$  variables, y está sujeta a una serie de restricciones, usaremos el método de los multiplicadores de Lagrange. El método de Lagrange soluciona el problema de las restricciones, sustituyendo las  $z$  restricciones por  $z$  variables, convirtiendo así el problema en un caso de  $n + z$  variables. Estas  $z$  nuevas variables reciben el nombre de multiplicadores de Lagrange  $\lambda_j \quad j=1, \dots, z$ . Una vez solucionado el problema de las restricciones, solo quedaría construir la función de Lagrange, y derivar esta función con respecto a las variables independientes e igualar a 0, hallando de este modo las condiciones de la función.

$$FL = \frac{\|w\|^2}{2} - \sum_{i=1}^N \lambda_j [y_i(w \cdot x_i + b) - 1] \quad (4)$$

$$\frac{\partial}{\partial w} \cdot FL = 0 \quad \frac{\partial}{\partial b} \cdot FL = 0 \quad (5)$$

Derivando y resolviendo las ecuaciones 5, los resultados quedan de la siguiente forma : [4]

$$w - \sum_{i=1}^N \lambda_j [y_i(x_i)] = 0 \quad \rightarrow \quad w = \sum_{i=1}^N \lambda_j [y_i(x_i)] \quad (6)$$

$$\sum_{i=1}^N \lambda_j [y_i] = 0 \quad (7)$$

Se puede apreciar que el vector  $w$  queda caracterizado mediante las ecuaciones anteriores, y por consiguiente el hiperplano también queda definido como una combinación lineal de los multiplicadores de Lagrange ( $\lambda$ ) para (i) distinto de 0.

Cuando los datos que se tratan no son separables y las fronteras de decisión no pueden ser lineales, se trasladan los datos de un espacio dimensional  $r$  a un espacio dimensional  $d$ , aplicando lo que se conoce como funciones matemáticas Kernel. Las funciones kernel representan un producto de puntos en un espacio dimensionalmente mayor: [2]

$$k(x_i, x_j) = x_i \cdot x_j = \phi(x_i) \cdot \phi(x_j) \quad i, j : 1, \dots, N \quad (8)$$

Una vez definida la función kernel se seguiría el modelo explicado en la parte superior.

### II-E. COMBINACIÓN DE DATOS

Las imágenes PET y las imágenes MRI proporcionan diferentes enfoques e información de una reconstrucción cerebral de un sujeto. La combinación de diferentes modelos de información proporciona una visión de estudio importante para detectar ciertas enfermedades y ayudar al diagnóstico médico. Sin embargo, para evaluar los resultados de diferentes fuentes de información, tenemos que tener en cuenta la etapa en la que los datos son integrados a lo largo del proceso de aprendizaje automático. De acuerdo con este enfoque, podemos hacer una integración temprana, intermedia y tardía. [6] [3]

**Integración temprana (Early integration).** La información de ambas fuentes es combinada antes del paso de clasificación en un solo vector de características. Los datos pueden simplemente concatenarse o someterse a una serie de transformaciones para reducir su dimensionalidad. Específicamente, durante el proyecto se construye un vector de características concatenando las características de las imágenes PET y las imágenes MRI:

$$V_i = (c_{i1}, c_{i2}, \dots, c_{im}, b_{i1}, b_{i2}, \dots, b_{in}) \quad (9)$$

donde  $(c_{i1}, c_{i2}, \dots, c_{im})$  y  $(b_{i1}, b_{i2}, \dots, b_{in})$  son, respectivamente, el vector de características de las imágenes PET y las imágenes MRI.

**Integración intermedia (Intermediate integration).** En este enfoque, también conocido como clasificación multicanal, la combinación se realiza dentro del propio clasificador utilizando dos matrices Kernel, una por cada fuente de datos. En lugar de hacer la combinación inicial, cogemos los kernel de ambos vectores de características y concatenamos los datos en función de estos kernel. Se pueden implementar enfoques lineales o no lineales. Durante el proyecto se implementa un kernel lineal, debido a las pequeñas bases de datos que se utilizan en el trabajo. La función de combinación :

$$K(x_i, x_j) = \sum_{m=1}^P W_m K_m(x_i^m, x_j^m) \quad (10)$$

donde  $P = 2$  siendo el número de kernel;  $w_m$  representan el peso de kernel  $k_m$ ;  $x_i, x_j$  son dos vectores de características y  $x_i^m, x_j^m$  son un subconjunto  $x_i, x_j$  con solo las características utilizadas para el kernel  $k_m$ .

**Integración tardía (Late integration).** Método de combinación de datos en el cual se entrena cada fuente de datos mediante un modelo de aprendizaje automático por separado y finalmente las predicciones de los diferentes modelos se combinan para tomar una decisión final. Esta combinación es realizada considerando la confianza de cada estimación. En la clasificación SVM la confianza de las estimaciones se calcula mediante la distancia de los datos de cada clase al hiperplano de margen máximo. Una vez que las fuentes de datos son clasificadas individualmente, se calculan las distancias de las diferentes clases al hiperplano, siendo estas  $ds$  y  $df$ . Si ambos clasificadores predicen la misma clase, esta clase es la elegida como clase correcta. En caso contrario, se calcula a través de las distancias al hiperplano ( $ds, df$ ) que

clase tiene mayor distancia y esta se asigna como la correcta. Como caso muy poco probable, donde las distancias de ambas clases son iguales, la clase correspondiente se asigna haciendo uso de la clasificación con mayor precisión en experimentos individuales.

### III. IMPLEMENTACIÓN

El desarrollo de la propuesta se basa en la realización de dos clasificaciones binarias para poder ayudar al diagnóstico entre sanos, enfermos o personas con deterioro cognitivo leve :

- La primera clasificación binaria realiza entre sujetos sanos y sujetos que tienen la AD. Para esta clasificación comparamos las imágenes NOR frente a las imágenes enfermos\_AD.
- La segunda clasificación binaria se realiza entre sujetos con deterioro cognitivo leve y sujetos que tienen la EA. Para esta clasificación comparamos las imágenes MCI frente a las imágenes enfermos\_AD.

El esquema de la figura 2 muestra los pasos en los que se distribuye la etapa de clasificación. Podemos observar que para cada situación se van a clasificar por un lado aquellas imágenes que son de tipo PET, por otro lado las de tipo MRI y finalmente se clasificará la combinación de ambos tipos de imagen PET+MRI. La etapa de clasificación se divide en diferentes sub-etapas:

- Clasificación con todos los vóxeles.
- Clasificación aplicando selección de características.
- Clasificación aplicando extracción de características (PCA).

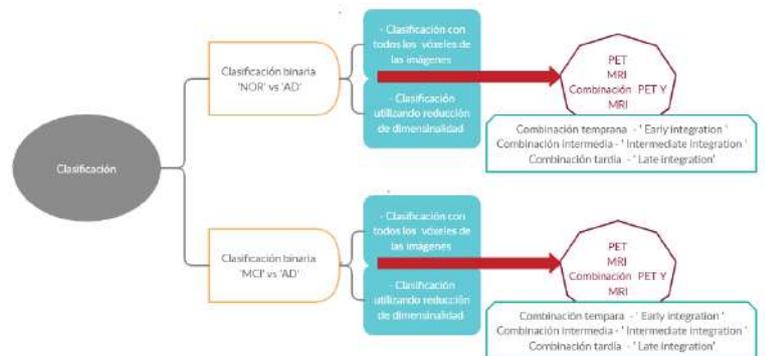


Figura 2. Implementación de la etapa de clasificación .

La elección del número de características se realiza mediante la utilización de diferentes criterios, definiendo el % de características que se van a utilizar del total de características. En el caso de utilizar selección de características la elección del número de características se realiza basándose en los valores absolutos de los resultados devueltos tras la métrica de la función 'rankfeatures'. En la utilización de PCA, la elección se realiza mediante la evaluación de la capacidad discriminante de cada autovector en función a su coeficiente de variabilidad (varianza de cada autovector) o a través de la evaluación de datos de

entrenamiento en una validación cruzada interna.

#### IV. RESULTADOS EXPERIMENTALES

A lo largo de esta sección se recogen y evalúan de manera clara y precisa los resultados de las etapas de clasificación indicadas en el desarrollo de la propuesta. Tras la implementación del código se analizan los resultados obtenidos, realizando un estudio sobre la influencia de la combinación de imágenes médicas en el diagnóstico de la AD.

Los resultados que se exponen a continuación, han sido obtenidos como la media de 5 procesos de clasificación diferentes, implementando 1 aleatorización distinta en cada proceso, y utilizando en cada clasificación validación cruzada.

Tenemos que destacar que durante la etapa de reducción de dimensionalidad vamos a tener 4 tipos distintos de combinación de imágenes :

- Early, Intermediate y Late .
- Union1, se basa en combinar las imágenes una vez que estas han sido procesadas en la etapa de reducción de sus características, es decir, más concretamente consiste en la unión de las características más importantes que se obtienen de las imágenes PET y la unión de las características más importantes que se obtiene de las imágenes MRI(gris).

#### IV-A. Comparación entre los clasificadores KNN, TREE y SVM, utilizando todas las características de las imágenes.

En primer lugar vamos a mostrar los resultados para los tres tipos de clasificadores que hemos desarrollado durante el proyecto. Los resultados se muestran para ambas clasificaciones binarias, observando en ellos el fundamento de la centralización del proyecto principalmente en la utilización de la clasificación SVM.

Precisión		NOR-enfermos_AD		MCI-enfermos_AD	
		media	desv	media	desv
SVM	PET	86,957	1,323	64,20	1,20
	MRI(ggris)	84,420	0,418	65,30	1,06
	EARLY	87,500	0,694	66,41	0,82
KNN	PET	72,029	1,503	61,44	1,48
	MRI(ggris)	63,333	2,268	60,22	0,87
	EARLY	72,609	0,324	60,11	1,48
TREE	PET	73,333	4,144	55,25	2,79
	MRI(ggris)	72,464	3,200	55,03	4,40
	EARLY	73,333	3,134	61,99	2,57

Cuadro I  
COMPARACIÓN CLASIFICADORES SVM, KNN Y TREE.  
(NOR-ENFERMOS\_AD)

Como podemos apreciar el cuadro I muestra los valores de rendimiento de los clasificadores SVM, KNN y TREE en la clasificación NOR frente enfermos\_AD y en la clasificación MCI frente enfermos\_AD . Centrándonos en los resultados podemos observar como el clasificador SVM proporciona resultados notablemente mejores, considerando con esto que la clasificación realizada por parte de SVM sería más fiable que la proporcionada por el resto de clasificadores a la hora de asignarlas etiquetas de predicción. Por esta razón durante este proyecto se implementa el clasificador SVM como modelo predictor.

#### IV-B. Resultados de la clasificación SVM para las distintas clasificaciones binarias.

Una vez que hemos elegido el clasificador SVM, vamos a realizar la clasificación de las dos clasificaciones binarias indicadas anteriormente. Estas clasificaciones se realizarán en diversas etapas dependiendo del número de características que se utilizan y la etapa en la que son combinadas.

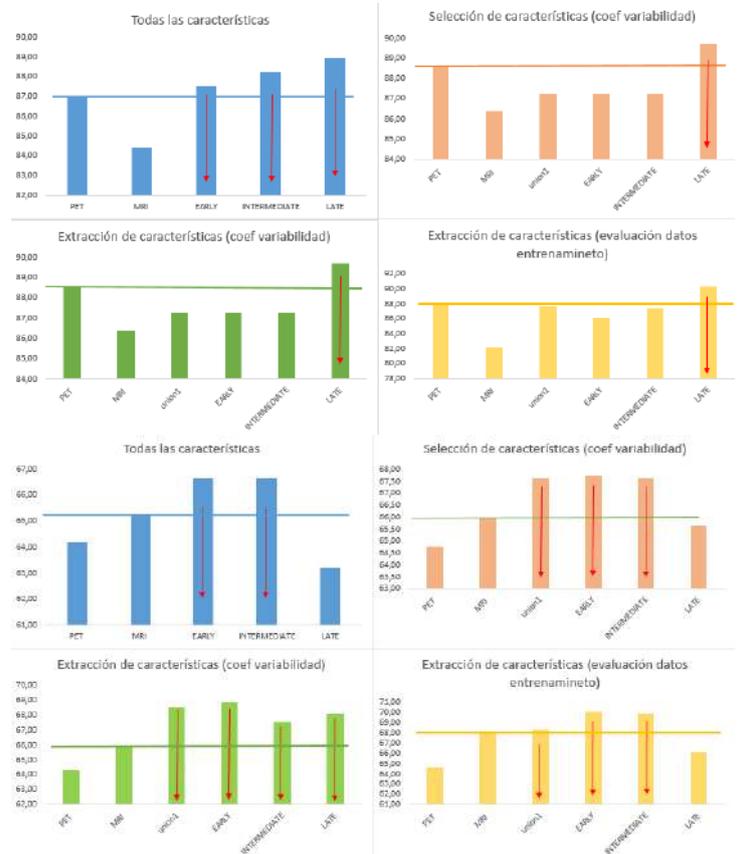


Figura 3. Representación gráfica de los rendimientos asociados a las dos clasificaciones binarias. Parte superior: PET - enfermos\_AD. Parte inferior: MRI - enfermos\_AD

La clasificación utilizando todas las características se basa en clasificar todos los vóxeles recogidos por las imágenes cerebrales sin discriminación en la importancia entre ellos. En el caso de realizar reducción de dimensionalidad podemos distinguir diversas formas de hacerlo a través de Pca o selección de características. Una vez que la función devuelve los índices de las características más importantes, se debe fijar un criterio para decidir que número de características es el más apropiado para la clasificación.

Los resultados de la imagen 3 y la tabla II muestran ciertas mejoras en la combinación de imágenes para las diferentes clasificaciones binarias. Los datos son analizados en profundidad en el apartado de conclusiones.

#### V. CONCLUSIÓN

En cuanto a los valores obtenidos durante todo el trabajo y basándonos en estudiar la combinación de diferentes modalidades de imagen médica cerebral para mejorar el diagnóstico

NOR-enferm_AD		Tod.caract	sel.caract	Ext.caract
Precisión media	PET	86,96	88,55	88,26
	MRI(gris)	84,42	86,38	83,33
	UnionI	—	87,25	87,92
	EARLY	87,50	87,25	86,38
	INTER	88,23	87,25	87,97
	LATE	88,95	89,71	90,31
Precisión máxima	PET	88,41	89,21	90,58
	MRI(gris)	84,71	86,96	85,51
	UnionI	—	87,68	88,41
	EARLY	88,41	89,13	88,41
	INTER	89,13	89,13	88,59
	LATE	89,85	92,03	91,83
MCI-enferm_AD		Tod.caract	sel.caract	Ext.caract
Precisión media	PET	64,20	64,75	64,31
	MRI(gris)	65,19	65,97	67,96
	UnionI	—	67,62	68,51
	EARLY	66,63	67,73	70,06
	INTER	66,63	67,62	69,89
	LATE	63,20	65,64	68,07
Precisión máxima	PET	65,75	67,40	69,71
	MRI(gris)	66,85	68,51	69,06
	UnionI	—	70,72	70,72
	EARLY	67,40	69,61	73,48
	INTER	66,85	70,17	70,31
	LATE	64,64	67,96	71,27

Cuadro II

RESUMEN DE LAS MEDIDAS DE RENDIMIENTO DE LA CLASIFICACIÓN SVM UTILIZANDO TODAS LAS CARACTERÍSTICAS, SELECCIÓN DE CARACTERÍSTICAS Y EXTRACCIÓN DE CARACTERÍSTICAS PARA LAS DOS CLASIFICACIONES BINARIAS.

asistido de la AD, la combinación de imágenes debemos indicar que no siempre mejora los valores de rendimiento. Lo que hacemos al unir las imágenes es aumentar los datos pero puede que la información no aumente y por tanto los valores de rendimiento no mejoren al combinar estas. Otro aspecto importante es que gran parte de los vóxeles de las imágenes aportan ruido y esta información no sirve para discriminar pacientes, sino que puede empeorar el proceso de predicción. Como conclusión podemos puntualizar en algunos aspectos:

- Para la primera clasificación binaria NOR frente enfermos\_AD, destacar 3 sucesos:
  - La combinación de imágenes utilizando todas las características mejora los valores de precisión de la clasificación.
  - La reducción de dimensionalidad mejora los valores de rendimiento de la clasificación.
  - La combinación de imágenes utilizando reducción de dimensionalidad no mejora los valores de precisión de la clasificación, a excepción de la late integration.
- Para la segunda clasificación binaria MCI frente enfermos\_AD, destacar 3 sucesos:
  - La combinación de imágenes utilizando todas las características mejora los valores de precisión de la clasificación, a excepción de la late integration.
  - La reducción de dimensionalidad mejora los valores de rendimiento de la clasificación.
  - La combinación de imágenes utilizando reducción de dimensionalidad mejora los valores de precisión de la clasificación, a excepción de la late integration.

Para la primera clasificación binaria NOR-enfermos\_AD los resultados que se obtienen aplicando reducción de dimensio-

nalidad son muy buenos llegando a alcanzar valores entorno al 90 % de precisión y por tanto los métodos de combinación en general no mejora la precisión. También destacar que para esta clasificación el método de combinación de imágenes que proporciona mejores resultados es la integración tardía donde las predicciones de los diferentes modelos se combinan para tomar una decisión final. En la segunda clasificación binaria, MCI-enfermos\_AD donde las imágenes pertenecientes a las diferentes clases no tienen tanta diferencia tanto funcionalmente como estructuralmente, la combinación de imagen aporta mejoras para el diagnóstico de la enfermedad, principalmente en la integración temprana. Finalizar indicando que la unión de datos es buena en situaciones concretas y no siempre unificar datos proporciona mejores resultados.

## REFERENCIAS

- [1] Gustavo A Betancourt. “Las máquinas de soporte vectorial (SVMs)”. En: *Scientia et technica* 1.27 (2005).
- [2] Shujun Huang y col. “Applications of support vector machine (SVM) learning in cancer genomics”. En: *Cancer Genomics-Proteomics* 15.1 (2018), págs. 41-51.
- [3] Andrea Mechelli y Sandra Vieira. *Machine Learning: Methods and Applications to Brain Disorders*. Academic Press, 2019.
- [4] Shahr Mendelson. *Advanced Lectures on Machine Learning: Machine Learning Summer School 2002, Canberra, Australia, February 11-22, 2002, Revised Lectures*. Vol. 2600. Springer Science & Business Media, 2003.
- [5] Jordi Peña-Casanova y Fundación la Caixa. “Enfermedad de Alzheimer: del diagnóstico a la terapia-conceptos y hechos”. En: (2007).
- [6] Fermín Segovia y col. “Combining PET images and neuropsychological test data for automatic diagnosis of Alzheimer’s disease”. En: *PLoS One* 9.2 (2014).
- [7] Emilio Corchado Hujun Yin y Vicente Botti Colin Fyfe. “Intelligent Data Engineering and Automated Learning—IDEAL 2006”. En: (2011).



**Sergio Ortiz Ruiz** nacido el 26 de Diciembre de 1998, en Granada. Graduado en Ingeniería de Tecnologías de Telecomunicación (2016-2020) por la Universidad de Granada..

# Extracción de información textural en neuroimagen

Autor: Miguel Bolívar Lupiáñez, e-mail: mboli96@correo.ugr.es

Tutores: Diego Salas González, e-mail: dsalas@ugr.es;

Javier Ramírez Pérez de Inestrosa, e-mail: javierrp@ugr.es;

Juan Manuel Górriz Sáez, e-mail: gorriz@ugr.es

Titulación: titulación, e.g. Ingeniería Informática o Ingeniería de Telecomunicación

**Departamento de Teoría de la Señal, Telemática y Comunicaciones**

**Universidad de Granada**

**Resumen**—El objetivo principal de este proyecto es introducir una nueva alternativa al estudio de imágenes cerebrales para enfermedades neurodegenerativas. El proyecto se basará en el uso de isolíneas como método para aislar regiones y obtener características físicas de las mismas que servirán como entrada de modelos de clasificación tales como máquinas de vectores de soporte. Para ello se dispone de una base de datos con imágenes SPECT de Alzheimer y control. Como entorno de desarrollo se utilizará Matlab, una herramienta muy optimizada en el uso de matrices como las que serán nuestras imágenes.

**Palabras clave**—Alzheimer, gradiente, isolíneas, máquinas de vectores soporte, morfometría basada en Voxel, rankfeatures.

## I. INTRODUCCIÓN

El Alzheimer es una enfermedad cerebral degenerativa que conduce a la pérdida de memoria, cambios en la forma de pensar o en la manera de comportarse. En España, actualmente hay alrededor de medio millón de pacientes con esta enfermedad. Se estima que para 2050 el número de pacientes haya aumentado hasta el millón. Este impactante dato nos concierne de la necesidad por investigar en este área. [1]

El Alzheimer en particular es una enfermedad del paciente y de su familia, porque requiere de un cuidador. Es una sobrecarga física, emocional y económica muy importante que afecta mucho a la vida de las personas.

Los avances tecnológicos en la informática actual permiten la obtención y manipulación de imágenes cerebrales. Estos avances abren una puerta al estudio de enfermedades neurodegenerativas, avanzando el diagnóstico ya que el juicio clínico se fundamenta en las neuroimágenes y en su procesamiento para definir distintas regiones cerebrales y analizar el estado en que se encuentran.

Con la motivación de colaborar en el desarrollo e investigación de esta enfermedad y con el conocimiento de procesamiento de imágenes, me dispongo a realizar este proyecto en el que analizaré con ayuda del entorno de desarrollo *Matlab* imágenes de pacientes sanos y enfermos de Alzheimer.

En primer lugar realizaré una introducción sobre los métodos de obtención de imágenes cerebrales y el método de estudio de las mismas más estandarizado, posteriormente fundamentaré el trabajo realizado en el que propongo un

método de estudio alternativo, concluyendo en los resultados y conclusiones.

### A. SPECT

La tomografía computarizada de emisión monofotónica es una técnica médica de tomografía que usa rayos gamma.[2]

La SPECT utiliza los rayos gamma que producen isótopos radioactivos como el tecnecio 99m. Estos isótopos se introducen en el cuerpo humano como parte de moléculas biológicamente activas. El procedimiento es similar al de la tomografía por emisión de positrones (PET), pero en la SPECT es el isótopo el que produce directamente el rayo gamma, mientras en la PET el isótopo produce un positrón que después se aniquila con un electrón para producir los dos rayos gamma. Estos dos rayos gamma salen en direcciones opuestas y su detección simultánea permite localizar el isótopo de forma más precisa que en la SPECT. La SPECT es, sin embargo, más simple porque pueden usarse isótopos más fáciles de obtener y de vida media más larga.

En general, se ha encontrado una buena correlación entre los hallazgos de SPECT y el juicio clínico. Por tanto es apta para medir funciones cerebrales regionales independientemente de los daños estructurales observados en la resonancia magnética.[3]

### B. Morfometría basada en Voxel

La Morfometría Basada en *Voxel* (VBM por sus siglas en inglés, *Voxel-Based Morphometry*) es una técnica de análisis en neuroimagen que permite la investigación de diferencias focales en la anatomía del cerebro, usando una aproximación estadística paramétrica. [3]

La VBM registra cada cerebro en una plantilla, la cual se deshace de la mayor parte de las grandes diferencias anatómicas del cerebro de las personas. A continuación, las imágenes cerebrales se suavizan para que cada *voxel* represente la media de sí mismo y sus vecinos. A partir de estas imágenes ya procesadas se realiza un análisis estadístico de grupos. El resultado de las pruebas de hipótesis realizadas para detectar dichas diferencias es un mapa estadístico paramétrico que muestra las regiones donde hay diferencias significativas entre unos cerebros y otros. [4]

## II. MÉTODOS

### A. Gradiente

En imágenes tridimensionales el gradiente de una función en un punto  $(x,y,z)$  viene dado por:

$$\nabla F = \frac{\partial F}{\partial x} \hat{i} + \frac{\partial F}{\partial y} \hat{j} + \frac{\partial F}{\partial z} \hat{k} \quad (1)$$

El gradiente mide la velocidad de cambio de una magnitud física a lo largo del espacio de la función. Para nosotros esa magnitud será el valor de intensidad del *voxel* que variará a lo largo de los tres ejes espaciales.

De esta manera la operación gradiente nos permite ampliar la base de datos obteniendo seis nuevas imágenes a partir de la original (gradiente en la dirección  $x$ ,  $y$ ,  $z$ , módulo del gradiente y ángulos de elevación y rotación).

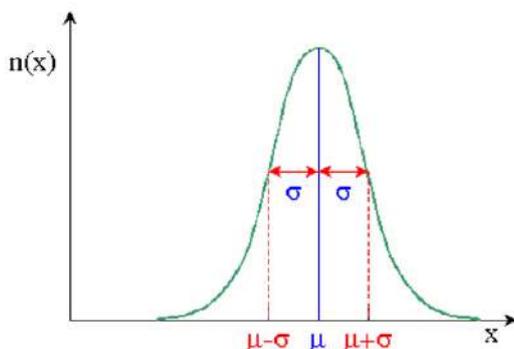
### B. Rankfeatures

Para realizar una clasificación de cerebros sanos y enfermos debemos hacer una comparación de los mismos en base a características físicas que posean. Para medir el grado de validez de una característica podemos utilizar RankFeatures que nos permite comprobar de manera cuantitativa qué características discriminan mejor un cerebro sano de uno enfermo. Para evaluar la separación entre clases existen dos criterios de evaluación.

1) *ttest*: este criterio considera las clases como variables aleatorias con distribución normal. Esta distribución está caracterizada por una media y una desviación típica como vemos en la siguiente ecuación [5]:

$$n(x) = \frac{n}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

La representación gráfica de este tipo de distribución tiene el siguiente aspecto:



Definiendo  $I$  como la característica que vamos a medir,  $I_\mu$  será la media de la misma,  $I_\sigma$  será la varianza de la característica, así como  $N$  será el número de muestras que tiene esa clase. De esta manera, el criterio de evaluación *ttest* calcula la separación entre clases de acuerdo a la siguiente expresión:[6]

$$Z = \left| \frac{I_{\mu\text{Clase1}} - I_{\mu\text{Clase2}}}{\sqrt{\left(\frac{I_{\sigma\text{Clase1}}}{N_{\text{clase1}}}\right)^2 + \left(\frac{I_{\sigma\text{Clase2}}}{N_{\text{clase2}}}\right)^2}} \right| \quad (3)$$

2) *wilcoxon*: Consiste en un test de clases no paramétrico. Esto significa que las características no siguen una regla de distribución como considerábamos con *ttest*. En este test se tomará un planteamiento de partida. En primer lugar, las observaciones de ambos grupos son independientes, además de ser variables ordinales o continuas. Bajo la hipótesis nula, la distribución de ambos grupos es la misma. Bajo otra hipótesis alternativa, los valores de una de las muestras tienden a exceder sobre los de la otra.[7]

### C. Máquinas de vector soporte

Las máquinas de vectores de soporte son un conjunto de algoritmos de aprendizaje supervisado pensados para resolver problemas de clasificación y regresión. A partir de un conjunto de entrenamiento etiquetados por clases se construye un modelo que sea capaz de predecir la clase de una muestra ajena al grupo de entrenamiento. [8]

La creación de este modelo surge a partir de la representación de los puntos en el espacio, separando las clases en dos espacios por un hiperplano definido como el vector entre los dos puntos de ambas clases más cercanos al que llamaremos vector soporte. De esta manera, nos interesa que las clases estén lo más separadas posibles. Para ello debemos buscar características lo más discriminatorias posibles.[9]

La función para clasificar un vector de características  $x$  se rige por la siguiente ecuación:

$$c = \sum_i \alpha_i k(s_i, x) + b \quad (4)$$

donde  $s_i$  son los vectores soporte,  $\alpha_i$  son los pesos,  $b$  es el *bias* y  $k$  una función de Kernel. Esta última resuelve el problema de la limitación de aprendizaje lineal de las máquinas.

### D. Isolíneas

Para delimitar zonas del cerebro podemos calcular isolíneas. Estas no son más que la unión de puntos de igual intensidad. De esta manera, aunque tenemos un conjunto finito de puntos en la imagen (*voxels*) esta técnica permite que la isolínea pase por cualquier punto del espacio, convirtiendo la imagen muestreada en una donde cualquier punto puede ser válido. Para saber sobre qué punto del espacio pasa una isolínea se utiliza una interpolación bilineal. [10]

v

## III. PROCEDIMIENTO

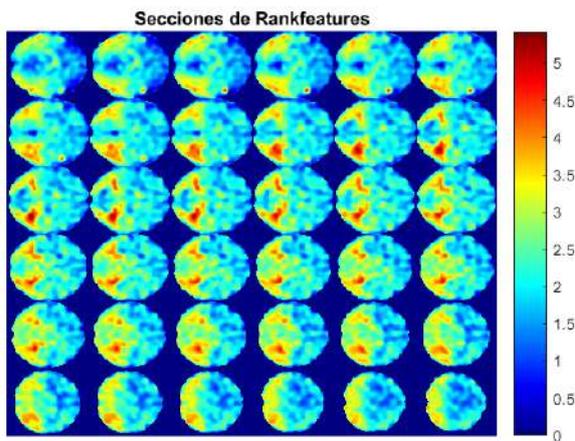
Se trabajará con la base de datos sin procesar y con la base de datos módulo del gradiente.

### A. Base de datos sin procesar

El objetivo es construir una nueva imagen resultado de haber aplicado la operación *rankfeatures* a toda la base de datos. El objetivo de este paso es tratar cada *voxel* como si de

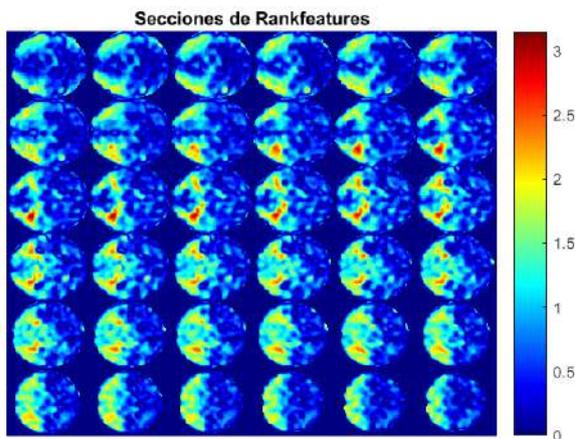
una característica se tratase. Si ese *voxel* tiene una intensidad muy parecida en todas las imágenes de la base de datos, la separación será muy pequeña y la función *rankfeatures* nos devolverá un valor muy bajo. Por el contrario, habrá *voxels* que tomen valores distintos en cerebros con Alzheimer y sanos, éstos darán valores de *rankfeatures* altos y así seremos capaces de saber de forma muy rápida qué zonas del cerebro debemos considerar para diferenciar la enfermedad de la salud.

1) *ttest*: Con el criterio de evaluación *ttest* representamos una serie de cortes transversales del cerebro para localizar las zonas de mayor relevancia. Representamos entre el corte 17 y 52:



Observamos que aproximadamente entre los cortes 28 y 32 se encuentran las zonas más rojizas del cerebro.

2) *wilcoxon*: : Realizamos el mismo procedimiento para el otro criterio de evaluación:

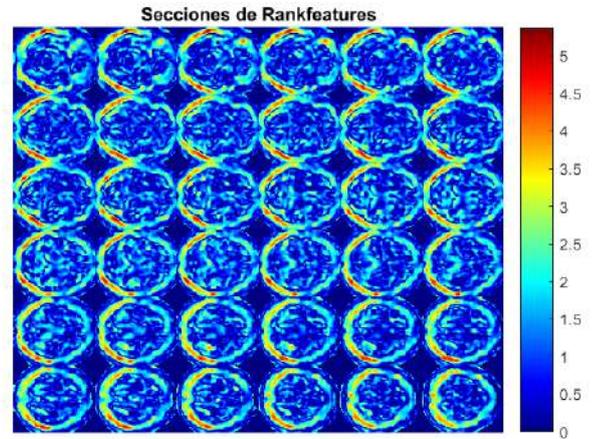


Vemos que ahora las secciones más relevantes se encuentran alrededor del corte 30.

### B. Base de datos gradiente

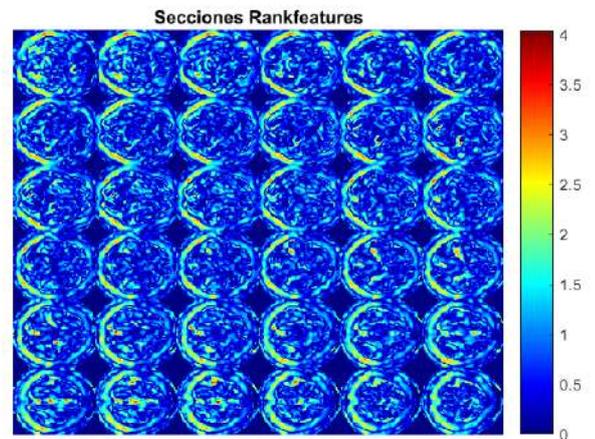
A continuación realizamos el mismo procedimiento para la base de datos gradiente.

1) *ttest*:



Vemos que los cortes más relevantes se encuentran sobre el corte 20.

2) *wilcoxon*:



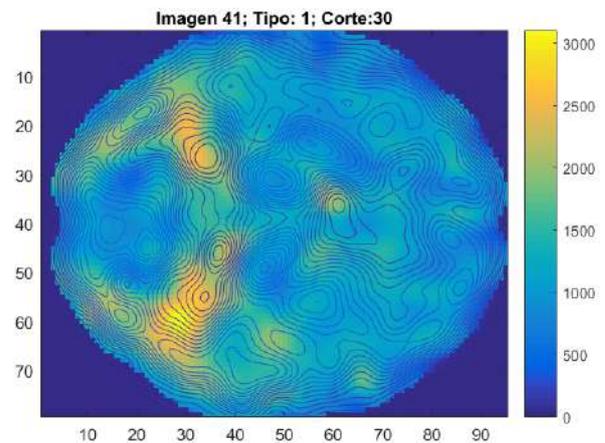
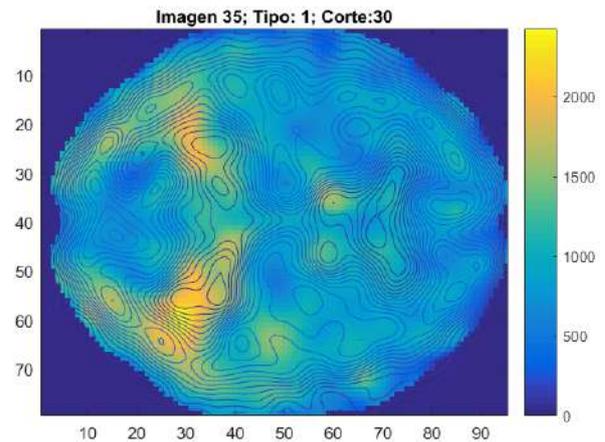
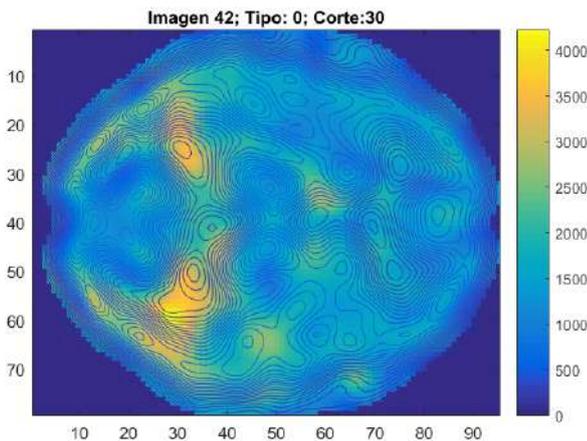
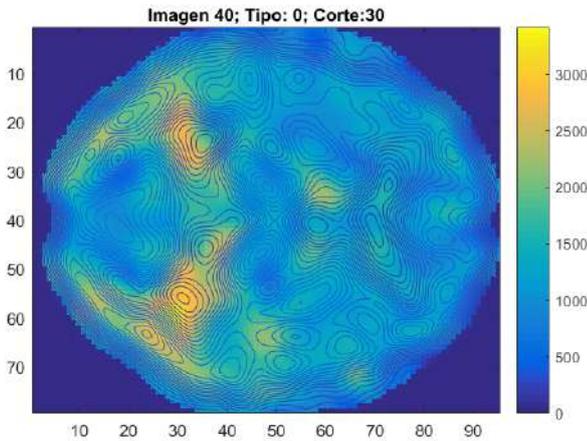
Igualmente al criterio anterior, los cortes más relevantes se encuentran en torno al 20.

### IV. ISOLÍNEAS CON IMÁGENES RANKFEATURES

Se superpondrán las isolíneas de algún corte con la imagen *rankfeatures* calculada antes. De esta manera podemos ver qué ocurre con las isolíneas en las zonas más relevantes del cerebro. Se escoge el corte 30 para la base de datos sin procesar.

A. Pacientes de control

Mostraremos dos imágenes de pacientes sanos.



Observamos un alto número de isolíneas en las zonas más amarillentas, así como un alto valor de intensidad en estas zonas. En este caso, la intensidad sí corresponde al valor de la imagen original en vez al valor de separación de clases. Si bien puede no equivaler al valor de los *voxels* exactamente, es al valor de interpolación entre varios.

Observamos dos características inmediatas que diferencian los cerebros sanos de los enfermos. La primera es el nivel de intensidad que toman los *voxels*. Este valor es mayor para pacientes sanos que enfermos. Es lógico ya que SPECT mide el nivel de flujo sanguíneo que corre por el cerebro y este, es mayor en pacientes sanos. La otra característica es el número de isolíneas que atraviesan las zonas determinantes. Este número es superior en pacientes sanos de enfermos.

En realidad, ambas características pueden estar relacionadas. Observando cualquiera de las imágenes *rankfeatures*, concluimos que existen dos tipos de zonas en el cerebro, unas determinantes y otras que no lo son. Ambas zonas se encuentran espacialmente distribuidas por los mismos puntos, de esta manera, la transición entre zonas determinante y no determinante también estará localizada en la misma zona para todos los cerebros. Aquellos cerebros cuyos *voxels* tengan más intensidad en la zona determinante tendrán una tendencia de cambio mayor que aquéllos cuyos *voxels* tengan valores de intensidad menos distante entre zonas determinantes y no determinantes.

B. Pacientes con Alzheimer

Igualmente mostraremos dos imágenes de pacientes enfermos.

V. CONCLUSIONES Y VÍAS FUTURAS

La mayor ventaja que obtenemos con la utilización de isolíneas es olvidarnos del concepto *voxel*. Ahora cualquier





# Segmentación de imágenes PET y SPECT usando Campos Aleatorios Ocultos de Markov

Autor: José Alberto Gómez López, e-mail: ajosegomez10@correo.ugr.es

Tutor: Fermín Segovia Román, e-mail: fsegovia@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—En los últimos años, se han propuesto varios sistemas de diagnóstico asistido por computadora para trastornos neurodegenerativos. Una gran rama de estos sistemas analiza datos de neuroimagen utilizando algoritmos de aprendizaje automático para estimar el diagnóstico de un determinado paciente. En este trabajo, se demuestra un método para seleccionar regiones de interés de las neuroimágenes PET y SPECT y, de ese modo, reducir su dimensionalidad. Se basa en Hidden Markov Random Field (HMRF). Primero, se calcula la media de todas las imágenes de una modalidad determinada. Posteriormente, esta imagen se segmenta utilizando un método de segmentación basado en HMRF. Como resultado, se obtienen diferentes mapas según la intensidad y la vecindad de los vóxeles (es más probable que un vóxel dado pertenezca al mismo mapa que los vóxeles de su vecindad). Los mapas resultantes se usan como una máscara para segmentar todas nuestras neuroimágenes. Posteriormente, se seleccionan los vóxeles correspondientes a los mapas más discriminativos. La capacidad discriminativa se estima mediante la distancia *Bhattacharyya*. El enfoque propuesto se evalúa utilizando dos conjuntos de datos de neuroimagen, uno con pacientes de Alzheimer y sujetos sanos y otro con pacientes y controles parkinsonianos. Después de la selección de la región, se usa un clasificador estadístico, en este caso la Máquina de Vectores de Soporte, para separar los controles y los pacientes utilizando las regiones seleccionadas. El rendimiento de la clasificación se estima mediante validación cruzada y se compara con el obtenido por los enfoques clásicos.

**Palabras clave**—Alzheimer, Parkinson, enfermedades neurodegenerativas, HMRF, regiones de interés, segmentación, normalización, aprendizaje automático, validación cruzada.

## I. INTRODUCCIÓN

La adquisición y análisis de datos de neuroimagen se ha convertido en un procedimiento común en la evaluación clínica y el seguimiento de trastornos neurodegenerativos. Las neuroimágenes cerebrales estructurales y moleculares se utilizan con frecuencia para ayudar al diagnóstico de la Enfermedad de Alzheimer (EA) y Parkinson (PD). Durante la última década, se han propuesto varios sistemas de Diagnóstico Asistido por Computadora, más conocido por sus siglas en inglés CAD, (*Computer-Aided Design*), que analizan y clasifican estas imágenes con alta precisión. Los algoritmos modernos de aprendizaje automático nos han permitido desarrollar CAD totalmente automáticos cuyas decisiones se basan únicamente en datos de neuroimagen. El enfoque clásico para clasificar los datos de neuroimagen utiliza la intensidad de todos los vóxeles cerebrales como característica, lo que conduce al pequeño problema del tamaño de la muestra: el número de observaciones (neuroimágenes) es muy reducido en comparación con la dimensionalidad

del problema (característica por neuroimagen) [1]. Se han propuesto varios enfoques para abordar este problema. Se pueden dividir en dos grupos según se requiera o no conocimiento previo. Por un lado, algunos sistemas realizan una selección de vóxeles en función de las regiones que se sabe que se ven afectadas por el trastorno (conocimiento previo). Por otro lado, los sistemas que no requieren conocimientos previos a menudo llevan a cabo una reducción basada en los algoritmos clásicos [2].

En este trabajo, se propone un método para reducir la dimensionalidad de los datos de neuroimagen a fin de abordar el problema del pequeño tamaño de la muestra y, de ese modo, mejorar el rendimiento de los sistemas CAD. Nuestra propuesta se basa en la selección de algunas regiones de interés y no requiere conocimientos previos sobre la enfermedad. Esta selección de región se basa en Campos Aleatorios Ocultos de Markov, en inglés *Hidden Markov Random Field (HMRF)*, una técnica clásica utilizada con éxito para segmentar datos estructurales [3]. Proponemos aplicar esta metodología para segmentar neuroimágenes moleculares como las obtenidas con Tomografía por Emisión de Positrones (PET) o Tomografía Computarizada por Emisión Monofotónica (SPECT).

## II. MATERIALES

Dos bases de datos se han utilizado para la evaluación de nuestra propuesta:

- **Base de datos para EA.** Esta base de datos está compuesta por 97 imágenes  $^{99m}\text{Tc}$ -ECD-SPECT de pacientes de Alzheimer y sujetos sanos. Las neuroimágenes fueron etiquetadas por tres clínicos experimentados después de analizar visualmente los datos. Como resultado, 41 imágenes fueron etiquetadas como 'normal', 56 como 'AD'. La mayoría de los pacientes en el último grupo (30) tienen EA en etapa temprana.
- **Base de datos para EP.** Esta base de datos contiene 189 imágenes  $^{123m}\text{I}$ -FP-CIT-SPECT de pacientes y síndromes parkinsonianos (SP). Después de analizar visualmente los datos, tres especialistas en medicina nuclear los etiquetaron, lo que dio como resultado 95 neuroimágenes etiquetadas como 'normal' y 94 como 'SP'. Aunque la EP es la patología más representativa del grupo de SP, esta base de datos contiene otras patologías con síntomas similares.

### III. PREPROCESADO DE LAS IMÁGENES

Una vez que todas las imágenes han sido adquiridas, resulta difícil poder compararlas entre sí porque pueden diferir en tamaño, en la distribución de valores de intensidad o en la orientación a la hora de la toma de las imágenes. Por estos motivos, es importante que se sigan unos pasos para hacer comparables todos los sujetos.

#### III-A. Normalización espacial

La complejidad de la estructura interna del cerebro, al igual que las diferencias entre cerebros de distintos pacientes, imposibilita la comparación entre ellos, por lo que se hace necesario un proceso de normalización espacial con respecto a una plantilla común. Este proceso nos permite comparar las imágenes cerebrales de los pacientes de una manera muy exhaustiva, vóxel a vóxel.

El proceso de normalización asume un modelo común con 12 parámetros [4] y una función de coste que presenta un valor óptimo cuando la plantilla utilizada y la imagen a normalizar se encuentran en la misma posición anatómica. La función objetivo a optimizar es habitualmente la diferencia cuadrática media entre ambas imágenes, la plantilla y la del paciente:

$$CF = \sum_i (f(Mx_i) - g(x_i))^2$$

donde  $f$  es la imagen a normalizar y  $g$  la plantilla utilizada. Para cada vóxel  $x = (x_1, x_2, x_3)$  de una imagen, la transformación afín a las coordenadas  $y = (y_1, y_2, y_3)$  se expresa mediante el producto matricial  $\mathbf{y} = \mathbf{M}\mathbf{x}$ .

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \quad (1)$$

Una vez se ha realizado la normalización, la imagen resultante se registra utilizando un modelo de transformación no rígido más complejo, ajustándose las deformaciones no lineales a la plantilla utilizada. Este proceso lo podemos observar en la Fig. 1.

#### III-B. Normalización en intensidad

Una vez se ha realizado la normalización espacial, se debe llevar a cabo una normalización en intensidad. Este proceso elimina diferencias de intensidad globales (máximas, medias o de ciertas regiones de referencia) entre pacientes. Estas diferencias pueden deberse al uso de diferentes cámaras de captación, a pequeñas diferencias en la dosis de radiofármaco que se inyecta a cada paciente o a como reaccione cada paciente frente al radiofármaco. Su aplicación es especialmente importante cuando se comparan imágenes adquiridas con cámaras diferentes. Se han estudiado tres propuestas de normalización en intensidad:

- **Normalización al máximo.** La primera propuesta consiste en dividir cada uno de los vóxeles de una imagen entre el valor máximo,  $I_{max}$ , que se obtiene como la media del 1% de la intensidad de los vóxeles de mayor intensidad.

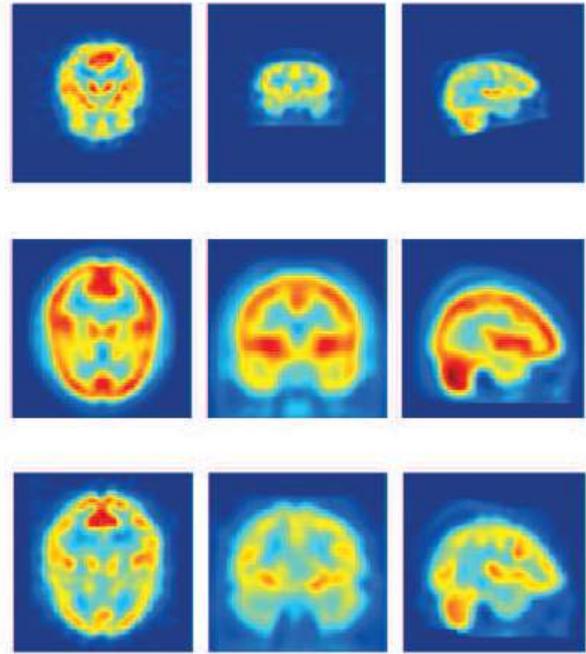


Figura 1. Normalización espacial de imágenes SPECT. Primera fila: Cortes originales. Segunda fila: Cortes de la plantilla. Tercera fila: Cortes normalizados [5].

- **Normalización a la media global.** Otra propuesta interesante es aquella que divide todos los vóxeles de una imagen entre el valor de intensidad medio de la propia imagen.
- **Normalización a una región no afectada.** Por último, y no de esta forma menos interesante, la última propuesta de normalización divide todos los vóxeles entre el valor de intensidad medio de una región concreta de la imagen. La región utilizada en nuestro caso es el cerebelo puesto que es una región no afectada por EA.

### IV. DESARROLLO DE LA PROPUESTA

Una vez realizado el preprocesado en todas las imágenes, se deben segmentar en regiones. Hay que destacar que el algoritmo de K-means se utiliza para inicializar el algoritmo HMRF que se aplicará posteriormente.

#### IV-A. Algoritmo K-means

*K-means* trata de clasificar un conjunto de muestras en varios subconjuntos en función de la semejanza de los atributos que describen cada subconjunto.

El procedimiento del algoritmo se basa en buscar un número de centroides  $\mu_k$  que sea igual al número de subconjuntos y asignar cada una de las muestras al centroide al que más se asemejen sus características. Un centroide representa el promedio de los puntos pertenecientes a una zona y se define de la siguiente manera:

$$\mu_k = (\mu_{k_1}, \mu_{k_2})^T \quad (2)$$

El conjunto de centroides se inicializan de manera aleatoria. Tras esto, se desarrolla un proceso iterativo en el que cada iteración se compone de dos pasos:

1. **Asignación de centroides:** Para cada muestra del conjunto se mide la distancia euclídea a cada centroide y se asigna al subconjunto cuyo centroide sea el que menor distancia tenga. Matemáticamente se ve de la siguiente manera:

$$c^{(i)} = \underset{k}{\text{máx}} \|x^{(i)} - \mu_k\| \quad (3)$$

2. **Movimiento de centroides:** Tras asignar los centroides, se calculan los nuevos como el promedio de todas las muestras asociadas a cada uno de ellos. Este paso se representa así:

$$\mu_k = \frac{1}{N_k} \sum_{x^i \in C_k} x^{(i)} \quad (4)$$

Una vez repetido el proceso iterativo un cierto número de veces, observaremos como se ha llevado a cabo una clasificación de todas las muestras en diferentes subconjuntos.

#### IV-B. Segmentación mediante HMRF

La segmentación basada en HMRF [6] asigna una etiqueta  $l_i \in L = \{1, \dots, M\}$ ,  $i = \{1, \dots, N\}$  a cada vóxel en una neuroimagen molecular según intensidad y vecindad. Por lo tanto, este procedimiento da como resultado la división de los vóxeles de una neuroimagen en  $M$  regiones.

Sean  $\mathbf{y} = \{y_1, y_2, \dots, y_N\}$  las ubicaciones de los  $N$  vóxeles que forman una neuroimagen. Cada uno de ellos está asociado a un valor de intensidad  $d_i$ . En este procedimiento, buscamos un etiquetado,  $\mathbf{x} = (x_1, x_2, \dots, x_N)$ , donde  $x_i \in L$  es la etiqueta asignada al vóxel en  $y_i$ . La estimación de  $\mathbf{x}$  se obtiene mediante el criterio MAP:

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \chi}{\text{arg max}} \{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})\} \quad (5)$$

donde  $\hat{\mathbf{x}}$  es una estimación de  $\mathbf{x}$  y considerada una realización particular del Campo Aleatorio de Markov  $X$  [6]. Usando la equivalencia entre los Campos Aleatorios de Markov y las distribuciones de Gibbs [3], la ecuación 5 se puede escribir como:

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \chi}{\text{arg min}} \{U(\mathbf{y}|\mathbf{x}) + U(\mathbf{x})\} \quad (6)$$

donde  $U(\mathbf{y}|\mathbf{x})$  es la energía de verosimilitud. Estimar  $\hat{\mathbf{x}}$  implica estimar el conjunto de parámetros  $\theta = \{\theta_l, l \in L\}$ , donde  $\theta_l = (\mu_l, \sigma_l)$ , ya que se asume que  $\mathbf{y}$  sigue una distribución gaussiana.

Por último, un algoritmo de Maximización de la Expectación estima alternativamente el conjunto de parámetros,  $\theta$ , y el conjunto de etiquetas,  $\hat{\mathbf{x}}$ .

En nuestra propuesta, la segmentación basada en HMRF se aplica una imagen (una por modalidad de datos) compuesta como la media de todas las neuroimágenes en un conjunto de datos determinado. Los mapas resultantes se usan como una máscara para segmentar todas las imágenes en el conjunto de datos correspondiente. La Fig. 2 muestra el resultado de la segmentación de las imágenes construidas en este trabajo,

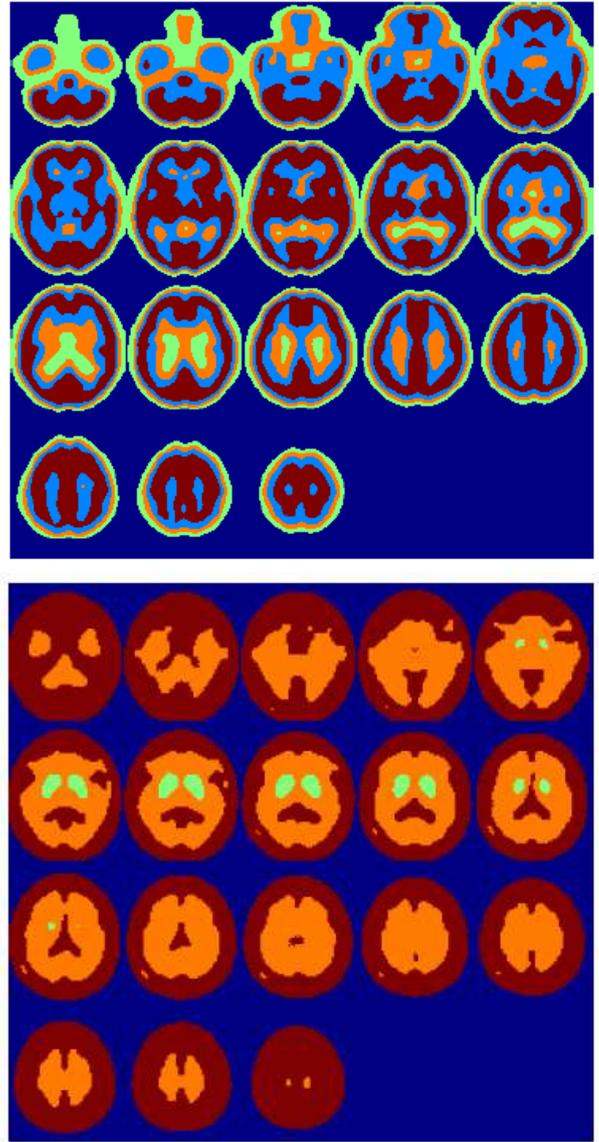


Figura 2. Maps resulting from the segmentation of the mean image of the ‘Dataset for AD’ (top) and from the segmentation of the mean image of the ‘Dataset for PD’ (bottom).

una correspondiente al “Conjunto de datos de AD” y otra correspondiente al “Conjunto de datos de PD”.

#### IV-C. Selección de características

Tras la segmentación basada en HMRF, se seleccionan aquellas regiones más discriminativas para cada una de las bases de datos (EA y EP). Para este fin, se calcula la distancia *Bhattacharyya* entre los grupos definidos en cada conjunto de datos utilizando cada región individual obtenida en la segmentación. Se seleccionan los vóxeles correspondientes a aquellas regiones que proporcionan las mayores distancias *Bhattacharyya* entre grupos. Esta distancia se obtiene de la siguiente manera:

$$B = \frac{1}{8} (\mu_1 - \mu_2)^T \left( \frac{\sigma_1 + \sigma_2}{2} \right)^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \ln \frac{|\frac{\sigma_1 + \sigma_2}{2}|}{\sqrt{|\sigma_1| |\sigma_2|}} \quad (7)$$

## V. CLASIFICACIÓN DE LOS SUJETOS

Después de haber aplicado nuestra propuesta a las diferentes bases de datos de las que se dispone, se hará una clasificación de los sujetos.

La mayoría de los sistemas de diagnóstico de enfermedades neurodegenerativas por ordenador utilizan un tipo especial de aprendizaje supervisado que proporciona etiquetas binarias, es decir, los objetos pueden ser etiquetados como: paciente sano y paciente enfermo. Tratando el diagnóstico de enfermedades neurodegenerativas, los objetos son presentados con vectores de características, cada uno de ellos construido a partir, o bien de imágenes funcionales del paciente o bien gracias a una imagen y otros datos del propio paciente.

El conjunto de datos con el que se va a trabajar en el proceso de clasificación se trata de un conjunto de vectores de características  $x_i \in \mathbb{R}^m, i = 1, \dots, n$ , siendo  $m$  la dimensión del espacio de características  $H$ . Este conjunto de datos utilizado para construir el clasificador se divide, a su vez, en 2 subconjuntos:

- **Datos de entrenamiento:** Aquel  $x_i$  que pertenece al subconjunto de entrenamiento  $X \subset \mathbb{R}^m$ . Matemáticamente:  $x_i \in X$ . Utilizado para construir la función de clasificación.
- **Datos de test:** Aquel  $x_j$  que pertenece al subconjunto de test  $Y \subset \mathbb{R}^m$ . Matemáticamente:  $x_j \in Y$ . Se usa para evaluar el funcionamiento de la función.
- La unión de estos dos subconjuntos formará el espacio de características  $H$ , es decir  $H = X \cup Y$ , asumiendo que  $H$  es  $\mathbb{R}^m$  [5].

Se debe cumplir que los conjuntos sean independientes entre sí. Esto significa que no pueden existir elementos del conjunto  $X$  (Datos de entrenamiento) en el conjunto  $Y$  (Datos de test).

$$X \cap Y = \{x_1, x_2, \dots, x_n\} \cap \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\} = \emptyset \quad (8)$$

El proceso de clasificación se divide en dos fases: etapa de entrenamiento y etapa de test o validación. Durante el entrenamiento se define un clasificador teniendo en cuenta los vectores de características de entrenamiento  $x_i \in X$ , cuyas etiquetas se conocen. El modelo exacto del clasificador dependerá en cierta medida del modelo que se haya propuesto para la clasificación.

La principal característica que se busca en un clasificador es que sea capaz de operar de manera satisfactoria con nuevos datos. Esta característica se denomina generalización.

Una vez separados los datos en Datos de entrenamiento y Datos de test se procede a realizar una ronda de validación cruzada. Una ronda de validación cruzada trata de analizar el subconjunto de entrenamiento y validar el resultado en el subconjunto de test. Con el objetivo de reducir la variabilidad en la evaluación global de la generalización, es habitual realizar varias rondas de validación cruzada utilizando distintos subconjuntos para entrenamiento y para test. Finalmente, el resultado se consigue realizando la media

aritmética de cada una de las iteraciones.

En el trabajo se han utilizado varias técnicas.

- **Validación por submuestreo aleatorio.** En cada iteración se divide el espacio muestral en dos subconjuntos aleatorios. Cuenta con la desventaja de que algunas muestras nunca se incluyan en el conjunto de evaluación.
- **Validación k-fold.** La base de datos se divide en  $k$  subconjuntos, en cada iteración (hay  $k$  iteraciones) se utiliza un subconjunto para entrenamiento y los  $k-1$  restantes para validación.
- **Leave one out.** Es un caso particular de  $k$ -fold. Cada muestra se evalúa una a una.

## VI. EVALUACIÓN DEL CLASIFICADOR

El objetivo principal que tiene un clasificador es asignar la etiqueta correcta a un objeto definido por un vector de atributos. Cuando se trata de un clasificador binario (paciente sano o patrón positivo y paciente enfermo o patrón negativo), este puede cometer dos errores básicos: clasificar como enfermo un paciente que en realidad es sano y viceversa. Cuando se compara la etiqueta estimada con la original existen distintas posibilidades:

- **Positivo Verdadero (PV):** Ambas etiquetas coinciden en valor positivo.
- **Positivo Falso (PF):** La etiqueta de test da un valor positivo mientras que la etiqueta original tiene valor negativo.
- **Negativo Verdadero (NV):** La etiqueta de test da un valor negativo mientras que la etiqueta original tiene valor positivo.
- **Negativo Falso (NF):** Ambas etiquetas coinciden en valor negativo.

A la hora de evaluar un clasificador se tienen en cuenta una serie de parámetros medidos una vez comparamos las etiquetas de test con las etiquetas originales:

- **Sensibilidad.** Se trata de la capacidad del clasificador para hallar positivos verdaderos. Puede expresarse como:

$$\text{Sensibilidad} = \frac{PV}{PV + NF} \quad (9)$$

Un clasificador que tenga una sensibilidad de 100% es aquel que tiene la capacidad de clasificar correctamente todos los objetos que han sido etiquetados como positivos. En el caso de que un clasificador con sensibilidad alta de un resultado negativo, podemos descartar la enfermedad puesto que este es muy fiable [5].

- **Especificidad.** Se trata de la capacidad del clasificador para detectar negativos verdaderos. Se expresa como:

$$\text{Especificidad} = \frac{NV}{NV + PF} \quad (10)$$

Un clasificador con una especificidad de 100% es aquel clasificador capaz de clasificar correctamente todos los objetos etiquetados como negativos. En el caso de que un clasificador con sensibilidad alta de un resultado

positivo, podemos confirmar la enfermedad puesto que este es muy fiable, raramente produce un resultado positivo [5].

- **Precisión.** Se trata de la capacidad del clasificador para etiquetar correctamente todos los patrones, ya sean positivos o negativos. Se expresa como:

$$Precision = \frac{PV + NV}{PV + PF + NV + NF} \quad (11)$$

El clasificador deseable es aquel que tiene valores altos de sensibilidad, precisión y especificidad. Puede haber la posibilidad de que existan clasificadores con un valor alto en la sensibilidad mientras que el de la especificidad es muy bajo. Este tipo de clasificadores no tiene la capacidad de diferenciar entre clases.

## VII. ANÁLISIS DE RESULTADOS

### VII-A. Base de datos 1

La base de datos Virgen de la Victoria contiene 189 imágenes SPECT de pacientes que sufren de Parkinson.

		Precisión	Sensibilidad	Especificidad
Seis regiones	2 ROIs	93.33 %	92.31 %	94.38 %
	3 ROIs	94.44 %	94.38 %	94.51 %
Estriado	-	88.88 %	87.09 %	90.80 %

Cuadro I

BASE DE DATOS VIRGEN DE LA VICTORIA. COMPARACIÓN ENTRE PRECISIÓN, ESPECIFICIDAD Y SENSIBILIDAD OBTENIDA CON NUESTRA PROPUESTA Y CON LOS ENFOQUES CLÁSICOS.

Se muestra como, tras segmentar las imágenes en diferentes regiones y luego escogiendo las más significativas, todos los resultados mejoran el estado del arte, es decir, son superiores a escoger como región afectada el estriado. Cabe decir que segmentando las imágenes en seis regiones y escogiendo las tres más significativas los resultados son óptimos. Se utiliza un clasificador SVM y una validación cruzada de 10fold.

### VII-B. Base de datos 2

La base de datos ADNI contiene imágenes PET de pacientes que sufren de Alzheimer. Esta base de datos contiene 420 imágenes.

		Precisión	Sensibilidad	Especificidad
Siete regiones	3 ROIs	89.28 %	87.16 %	91.36 %
	4 ROIs	89.64 %	87.75 %	91.27 %
Todo el cerebro	-	89.28 %	86.66 %	92.31 %

Cuadro II

BASE DE DATOS ADNI. COMPARACIÓN ENTRE PRECISIÓN, ESPECIFICIDAD Y SENSIBILIDAD OBTENIDA CON NUESTRA PROPUESTA Y CON LOS ENFOQUES CLÁSICOS.

El Cuadro anterior muestra los resultados pertenecientes a pacientes enfermos de Alzheimer y controles. En este caso el estado del arte se mejora levemente, es decir, tanto la precisión como la sensibilidad son ligeramente mayores si las imágenes se segmentan en siete regiones y se escogen las tres o cuatro más significativas que utilizando todos los vóxeles, considerado estado del arte.

### VII-C. Base de datos 3

La base de datos Virgen de las Nieves contiene imágenes SPECT de pacientes que sufren de Alzheimer. Esta base de datos contiene 97 imágenes.

		Precisión	Sensibilidad	Especificidad
Siete regiones	3 ROIs	77.08 %	74.51 %	78.94 %
	4 ROIs	82.29 %	78.57 %	85.18 %
Todo el cerebro	-	81.25 %	76.74 %	84.90 %

Cuadro III

BASE DE DATOS VIRGEN DE LAS NIEVES. COMPARACIÓN ENTRE PRECISIÓN, ESPECIFICIDAD Y SENSIBILIDAD OBTENIDA CON NUESTRA PROPUESTA Y CON LOS ENFOQUES CLÁSICOS.

Este Cuadro nos muestra como hay casos en los que el estado de arte se mejora. En este caso, el estado del arte nos dice que el clasificador se evalúa utilizando todos los vóxeles y, de esta manera, obtenemos un 81 % de precisión. Sin embargo, utilizando las cuatro regiones más discriminativas cuando las imágenes se segmentan en siete regiones, la precisión es ligeramente superior a la del estado del arte.

## VIII. CONCLUSIONES

Tras analizar los resultados que se han obtenido en la sección anterior y teniendo en cuenta los objetivos que se propusieron al comienzo del trabajo, las conclusiones obtenidas son las siguientes:

- Se mejoran los resultados dados con los enfoques clásicos.
- El clasificador SVM obtiene los resultados óptimos ya que se adapta mejor a problemas con un tamaño muestral limitado y que tiene gran capacidad de generalización.
- El método de validación cruzada que ofrece mejores resultados es 10-fold puesto que se ve menos afectado a un posible sobreentrenamiento del clasificador al usar un grupo más amplio de test.

## IX. PROYECCIÓN DE FUTURO

Una vez vistos los buenos resultados obtenidos, en el futuro se pretende:

- Ampliar a otro tipo de enfermedades neurodegenerativas en las cuales no exista un método de diagnóstico precoz.
- Ampliar a otro tipo de imágenes médicas como pueden ser MRI o TAC.
- Utilizar otros métodos de segmentación para comparar.
- Estudiar la clasificación con redes neuronales convolucionales.

## AGRADECIMIENTOS

A mi abuelo Fernando, él estaría orgulloso.

## REFERENCIAS

- [1] Schnack, Hugo G. and Kahn, René S, "Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters", *Frontiers in Psychiatry*, vol. 7, Mar. 2016.
- [2] S. Rathore *et al.*, "A Review on Neuroimaging-Based Classification Studies and Associated Feature Extraction Methods for Alzheimer's Disease and Its Prodromal Stages", *Neuroimage*, vol. 155, no. Supplement C, pp. 530-548, Jul. 2017.

- [3] Y. Zhang *et al.*, “Segmentation of brain MR imagen through a hidden Markov random field model and the expectation-maximization algorithm”, *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45-47, Jan. 2001.
- [4] R. P. Woods, “Spatial transformation models”, *Handbook of medical imaging: processing and analysis*, pages 465-497.
- [5] F. Segovia Román, “Análisis de Imágenes Funcionales Cerebrales Mediante Modelos de Mezcla de Gaussianas y Mínimos Cuadrados Parciales para el Diagnóstico de Alteraciones Neurológicas”, *Universidad de Granada*.
- [6] Z. Li, *Markov Random Field Modeling in Image Analysis*, 2nd ed., ser. Computer Science Workbench. Tokyo, Japan: Springer, 2001.



**José Alberto Gómez López** . Nació el 07/06/1996 en Bélmez de la Moraleda, Jaén. Comenzó el Grado de Ingeniería de Telecomunicación en 2014/2015 en la Universidad de Granada, finalizándolo 2017/2018.

Cabe destacar la presentación de un póster en el congreso internacional NSS/MIC 2018. El congreso Nuclear Science Symposium (NSS) y Medical Imaging Conference (MIC) se llevará a cabo en el International Convention Centre Sydney, en Sydney, Australia, del 10 al 17 de noviembre de 2018.

# SEGMENTACIÓN DEL CUERPO ESTRIADO EN IMÁGENES FP-CIT SPECT PARA EL DIAGNÓSTICO DEL PARKINSON

Autor: José Miguel Mateos Ramos, e-mail: josemiteleco@correo.ugr.es

Tutor: Diego Salas González, e-mail: dsalas@ugr.es

Titulación: Ingeniería de Tecnologías de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones

Universidad de Granada

**Resumen**—Una de las enfermedades neurodegenerativas más comunes actualmente es la conocida como enfermedad de Parkinson. Por tanto, el objetivo del presente trabajo es la clasificación entre pacientes sanos y enfermos, haciendo uso de imágenes funcionales FP-CIT SPECT. En concreto, se analizarán una serie de volúmenes tridimensionales que rodean al cuerpo estriado, denominados isosuperficies.

La principal novedad de este estudio es la estimación automática de una isosuperficie umbral que contenga al cuerpo estriado. A partir de dicha isosuperficie, se extraen un conjunto de características que ayuden en el diagnóstico del Parkinson. Los resultados obtenidos proporcionan tasas de clasificación entorno al 90%. Además, su principal ventaja respecto a trabajos previos es que no deforma la imagen de entrada, manteniendo la interpretación física y visual de la misma.

**Palabras clave**—cuerpo estriado, isosuperficies, machine learning, neuroimagen, Parkinson, procesamiento de imágenes.

## I. INTRODUCCIÓN

LA enfermedad de Parkinson es un trastorno neurodegenerativo caracterizado por la pérdida de neuronas dopaminérgicas en la sustancia negra. Sus principales síntomas son temblores involuntarios, rigidez muscular, dificultad al andar, etc. En medicina, las imágenes funcionales FP-CIT SPECT han sido ampliamente utilizadas para el diagnóstico del Parkinson (PD), dado que son un indicador preciso de los transportadores dopaminérgicos (DAT) en el cerebro. Estas imágenes se producen mediante la inyección al paciente de una sustancia radioactiva conocida como  $I^{123}$ -Ioflupano, que se adhiere a los DATs del cerebro.

En un paciente sano, la concentración de DATs resulta en una imagen con altos valores de intensidad en aquellas zonas en la que la concentración del  $I^{123}$ -Ioflupano es mayor. Además, la forma que presenta el cuerpo estriado es curvada. Por el contrario, en los pacientes con PD suele apreciarse un estriado más redondeado y con valores de intensidad más bajos, lo que implica una reducción aparente del volumen del estriado.

El reciente uso de isosuperficies aplicado al Parkinson ha aportado una serie de marcadores físicamente interpretables que pueden ser muy útiles en diagnóstico de pacientes. Dada una imagen de un sujeto, dicha imagen se puede dividir en varias isosuperficies, de acuerdo con la intensidad de esta. Por tanto, las isosuperficies interiores están asociadas a valores

de intensidad más altos. Si analizamos estas isosuperficies desde el interior hacia el exterior, existe por lo general una isosuperficie que presenta un volumen mucho mayor que la isosuperficie anterior. Esta isosuperficie se puede emplear como frontera de la región más iluminada de la imagen. Con lo cual, nuestra hipótesis es que dicha isosuperficie (y las isosuperficies que contiene) presenta suficiente información para diferenciar a ambos tipos de sujetos.

## II. REVISIÓN DEL ESTADO DEL ARTE

Durante varios años, para abordar este tipo de problemas se empleaban transformaciones a la imágenes de entrada para adaptarlas a una plantilla, como se aprecia por ejemplo en la Figura 1, tomada de [1]. A partir de las imágenes transformadas, se comparaban los valores de intensidad de todas las imágenes de la base de datos para distinguir entre los distintos sujetos. Sin embargo, con esta metodología se pierde la interpretación física de las distintas regiones del cerebro para dar paso a una comparación voxel a voxel [2].

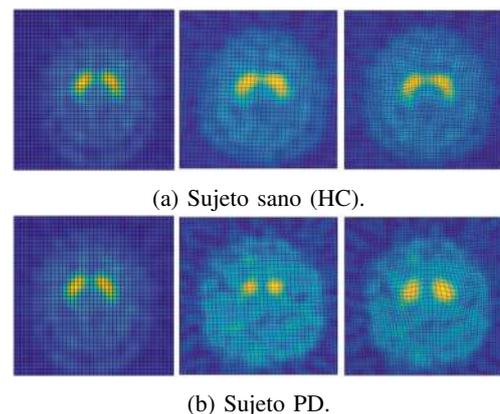


Fig. 1: Transformaciones afín y no afín aplicadas a imágenes HC y PD. En ambos casos a la izquierda se encuentra la plantilla objetivo de la transformación, en el centro el resultado de una transformación afín y a la derecha el resultado de una transformación no afín.

Por otra parte, en trabajos más recientes como [3], se emplearon todas las isosuperficies de una imagen para la clasificación, con resultados prometedores. El objetivo de este

trabajo es continuar con esta línea de investigación, de modo que en lugar de emplear todas las isosuperficies de la imagen indiscriminadamente, se empleen solo aquellas que afecten a la distinción de pacientes.

### III. PLANIFICACIÓN TEMPORAL

A modo de resumen, en la Figura 2 se ilustra un diagrama de Gantt con la planificación temporal destinada a este trabajo, con un total de 300 horas estimadas.

	Oct.	Nov.	Dic.	Ene.	Feb.	Mar.	Abr.	May.	Jun.
Estado del arte									
Procesado de imágenes									
Cálculo de $\alpha$ -shapes									
Parámetros asociados									
Delimitación del estriado									
Características del estriado									
Clasificador y resultados									
Redacción									

Fig. 2: Diagrama de Gantt con el resumen de la planificación temporal.

### IV. ESTIMACIÓN DE COSTES

En la Tabla I se incluye el desglose de todos los costes del proyecto. Dado que las imágenes de la base de datos necesitan ser tomadas en un hospital, se ha incluido el sueldo del médico asociado, atendiendo a los datos de [4].

Activo	Sueldo (€/hora)	Horas trabajo	Coste total (€)
Ingeniero	12.27	300	3681.00
Doctor	27.70	208	5761.60
Margen	-	-	944.26
<b>Total</b>			<b>10386.86</b>

Tabla I: Desglose de los costes asociados al proyecto.

### V. MATERIALES Y MÉTODOS

#### A. Base de datos

Las imágenes de las que consta nuestra base de datos han sido tomadas en el Hospital Virgen de la Victoria de Málaga (España). En total, se han tomado un conjunto de 208 imágenes, de las que 108 corresponden a sujetos sanos y 100 al grupo de pacientes etiquetados como PD. En la tabla II se reflejan los datos estadísticos de los pacientes, donde la edad se expresa de la forma  $\bar{m} \pm \sigma$ , con  $\bar{m}$  la media de la edad de cada grupo en concreto y  $\sigma$  la desviación estándar.

	Proporción		Edad		
	Hombres	Mujeres	Hombres	Mujeres	Ambos
HC	48.1%	51.9%	68.5 $\pm$ 11.0	69.9 $\pm$ 9.3	69.3 $\pm$ 10.2
PD	43.0%	57.0%	67.6 $\pm$ 10.8	68.8 $\pm$ 8.7	68.3 $\pm$ 9.6

Tabla II: Datos de los pacientes participantes en el estudio.

#### B. Procesado de imágenes

1) *Normalización en intensidad*: En primer lugar, al conjunto de las imágenes de la base de datos se le ha aplicado una normalización en intensidad basada en la distribución  $\alpha$ -estable. Esta normalización se aplica para que las diferencias entre distintas imágenes se deban realmente a características morfológicas, en lugar de a los parámetros de la cámara gamma usada en la adquisición de las imágenes o la cantidad del fármaco Ioflupano que se le ha inyectado a un paciente. A modo de ejemplo, en la Figura 3, perteneciente a [5], se muestran los histogramas de intensidad de una imagen antes y tras aplicar esta normalización. En el segundo caso, vemos como los distintos histogramas de todas las imágenes son mucho más parecidos.

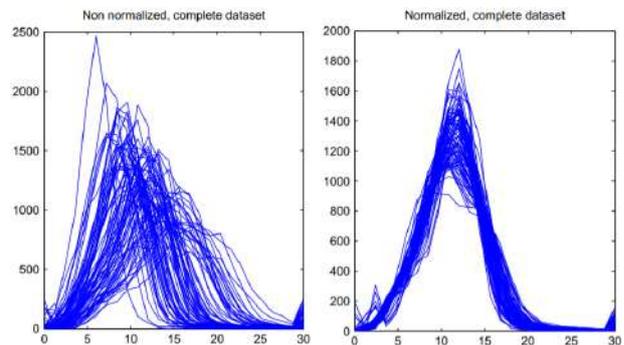


Fig. 3: Histogramas superpuestos de varias imágenes (izquierda) y los histogramas resultantes tras aplicar la normalización en intensidad (derecha).

2) *Isosuperficies*: Para hallar las distintas isosuperficies de una imagen, se han empleado las denominadas  $\alpha$ -shapes, que se tratan de volúmenes tridimensionales que contienen a una serie de puntos [6]. Cabe destacar que las isosuperficies pueden estar divididas en varias regiones que se encuentren separadas en el espacio, lo que será de importancia más adelante.

Sin embargo, al hallar las isosuperficies de una imagen, podemos encontrarlos con el inconveniente que se refleja en la Figura 4. En ella, se observa que aparecen regiones de alta intensidad fuera del cuerpo estriado, y que podrían dar lugar a errores para la obtención de la isosuperficie umbral.

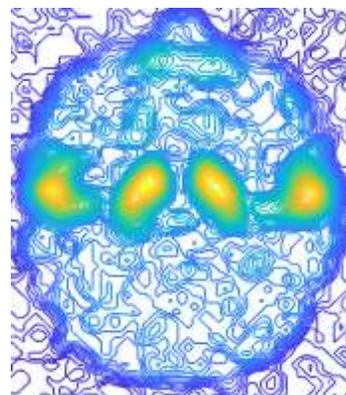


Fig. 4: Isolíneas correspondientes a un corte axial de la imagen de entrada.

Para solventar este problema, se ha hallado un paralelepípedo tridimensional en cada una de las imágenes, tal y como se refleja en la Figura 5.

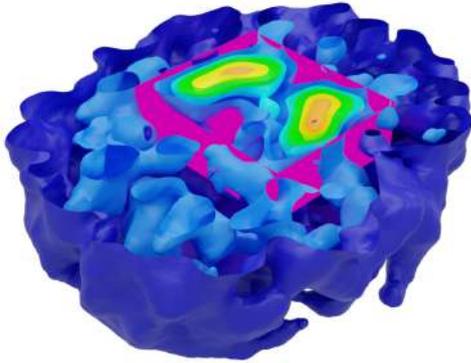


Fig. 5: Ejemplo de 7  $\alpha$ -shapes obtenidos de un sujeto sano. El volumen púrpura limita la región en la que potencialmente se encuentra el estriado.

Este paralelepípedo se calcula de manera dinámica para cada nueva imagen de entrada. El algoritmo parte del cálculo de 3 imágenes, halladas a partir de la imagen tridimensional. Las imágenes constan de los máximos valores de intensidad en cada uno de los planos: axial, sagital y coronal. En la Figura 6 se muestra un ejemplo de las imágenes en cada uno de los planos.

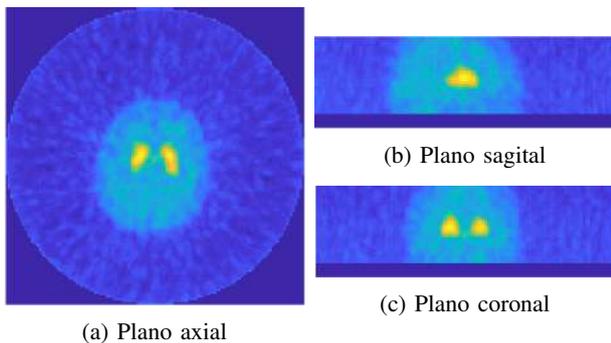


Fig. 6: Imágenes correspondientes al sujeto 1, obtenidas como los máximos valores de intensidad en cada uno de los planos.

Posteriormente, se hallan 3 señales unidimensionales a partir de las imágenes mencionadas previamente, empleando de nuevo el máximo valor de las imágenes en uno de los ejes. El resultado de este proceso se muestra en la Figura 7. En base a estas 3 gráficas se hallan los puntos marcados en rojo en las mismas, que representan los límites del paralelepípedo en cada uno de los ejes espaciales.

Una vez hallado dicho paralelepípedo, o mejor dicho, sus límites, se aplica el siguiente algoritmo para obtener las isosuperficies correspondientes a cada hemisferio del estriado:

- Se halla la isosuperficie de mayor intensidad cuyo centro se encuentra dentro del paralelepípedo. De aquí en adelante la denotaremos como  $L_1$  (Layer 1). El centro de una isosuperficie se calcula como la media de las coordenadas de los voxels que la conforman. La isosuperficie  $L_1$  debe estar formada por una sola región, al igual que para el resto de isosuperficies. Si el algoritmo

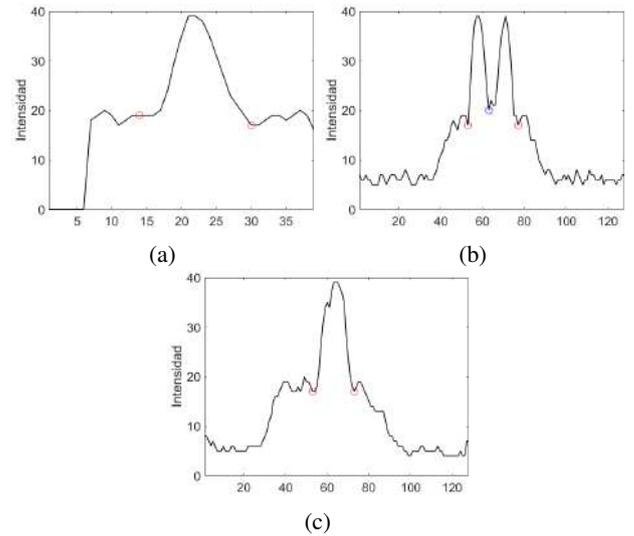


Fig. 7: Señales unidimensionales calculadas a partir de las imágenes bidimensionales del cerebro. En rojo se marcan los puntos que se considerarán como los límites del paralelepípedo. Con un círculo azul se rodea el punto medio de ambos hemisferios del cuerpo estriado.

del cálculo de los bordes del paralelepípedo es correcto,  $L_1$  tendrá una única región de alta intensidad dentro de dicho paralelepípedo.

- Para las demás isosuperficies, se escoge la región cuyo centro sea el más cercano a  $L_1$ . A dicha región,  $L_n$ , se le calcula su volumen y superficie. Sin embargo, si el centro de  $L_n$  se encuentra a más de 15 voxels de distancia de  $L_1$ , sus parámetros se ponen a 0 para evitar costes computacionales innecesarios. Empíricamente se ha hallado que  $L_T$  se encuentra a una distancia menor de 15 voxels de  $L_1$ .
- Se guardan los vectores de parámetros para cada imagen y el  $\alpha$ -shape que presenta 1 sola región en cada iteración.

### C. Regiones de interés

A partir del algoritmo anterior, para cada imagen, obtendremos 2 gráficas para cada hemisferio del estriado, correspondientes al volumen y a la superficie en función de la intensidad. En la Figura 8 podemos ver un ejemplo de ellas.

Si recordamos la hipótesis de la que partíamos, se asumía que a partir de una isosuperficie, el volumen de las isosuperficies crecía exponencialmente. También se asume que ocurre el mismo fenómeno para la superficie. Por lo tanto, a partir de gráficas como la Figura 8, procederemos del siguiente modo: primero se halla la derivada normalizada de uno de los 2 parámetros, por ejemplo volumen, del siguiente modo:

$$\Delta V(n) = 2 \frac{V(n) - V(n+1)}{V(n) + V(n+1)} \quad (1)$$

Posteriormente, volvemos a aplicar una derivada a  $\Delta V(n)$ :

$$\Delta^2 V(n) = \Delta V(n) - \Delta V(n+1) \quad (2)$$

Finalmente, la isosuperficie umbral ( $L_T$ ) se calcula como el último valor positivo de  $\Delta^2 V(n)$  por debajo de un cierto

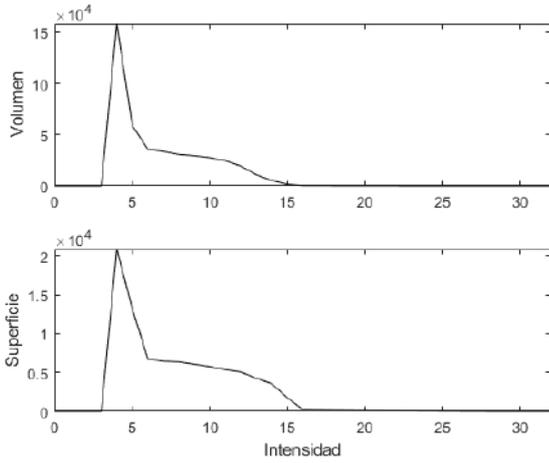


Fig. 8: Volumen (arriba) y superficie (abajo) de las isosuperficies del hemisferio izquierdo en función de la intensidad.

umbral  $t$ , más una unidad. Dicho umbral se ha establecido como el 95% de la intensidad máxima de la imagen.

$$L_T = \max \{n \in \mathbf{Z} \leq t : \Delta^2 V(n) > 0\} + 1 \quad (3)$$

#### D. Características de clasificación

Dada una isosuperficie umbral y las isosuperficies interiores, en la Tabla III se muestran todas las características que se han hallado para cada hemisferio de la imagen.

Hemisferio izquierdo	Hemisferio Derecho
Volumen de $L_T$	Volumen de $L_T$
Superficie de $L_T$	Superficie de $L_T$
Esféricidad de $L_T$	Esféricidad de $L_T$
Distancia a $L_1$	Distancia a $L_1$
Área de la diferencia de volúmenes	
Área de la diferencia de superficies	
Área de la diferencia de esfericidades	
Área de la diferencia de distancias a $L_1$	
Intensidad de $L_T$	Intensidad de $L_T$
Intensidad máxima	Intensidad máxima
Intensidad media	Intensidad media
Percentil 75	Percentil 75
Mínimo radio del $\alpha$ -shape	Mínimo radio del $\alpha$ -shape
Vectores de gradiente	Vectores de gradiente

Tabla III: Resumen de las características extraídas del cuerpo estriado.

De dichas características, cabe destacar que la esfericidad de un volumen mide el parecido de dicho volumen con una esfera, y se calcula como:

$$\Psi = \frac{(36\pi V_{obj})^{1/3}}{S_{obj}} \quad (4)$$

En cuanto al área de la diferencia de algunos parámetros, estas se calculan como:

- 1) Para cada isosuperficie de cada hemisferio del estriado, se calcula su volumen, superficie, esfericidad y distancia a  $L_1$ . Mediante este procedimiento obtendremos un conjunto de valores para cada característica.
- 2) Si el número de valores para una determinada característica es distinto en ambos hemisferios, se rellenará

con ceros por la derecha el conjunto más pequeño a fin de igualar el número de valores de todas las características en cada hemisferio.

- 3) Calculamos el valor absoluto de la diferencia de valores entre ambos hemisferios del estriado, para cada característica. De este proceso resultan 4 gráficas distintas, una para cada rasgo.
- 4) Hallamos el área bajo la curva de cada una de las 4 gráficas anteriores, desde  $L_{T_{max}}$  hasta el máximo valor de cada imagen, donde:

$$L_{T_{max}} = \max\{L_{T_{izq}}, L_{T_{der}}\} \quad (5)$$

Asimismo,  $L_{T_{izq}}$  y  $L_{T_{der}}$  son las isosuperficies umbrales de cada hemisferio del estriado.

El objetivo de este procedimiento es cuantificar las diferencias entre ambos hemisferios del estriado, de ahí que el área bajo cada curva se calcule a partir de  $L_{T_{max}}$ , puesto que de este modo se garantiza que únicamente se toman valores pertenecientes al cuerpo estriado.

Finalmente, el gradiente se trata de una discretización en 32 regiones espaciales de los ángulos de elevación y azimutal del gradiente, realizando la media de la magnitud del gradiente para aquellos vectores que pertenezcan a la misma región.

#### E. Clasificación basada en Machine Learning

Para clasificar a los distintos sujetos, se han empleado 2 técnicas de clasificación. La primera de ellas se trata de una máquina de vectores soporte con kernel lineal (y parámetro  $C = 1$ ). La otra alternativa ha sido un clasificador basado en Adaptive Boosting. Además, en ambos casos se ha aplicado el esquema de validación cruzada 10-fold.

Por otra parte, ya que no todas las características son relevantes en la clasificación, se han ordenado las características en función de 5 rankings: ttest, entropy, roc, Bhattacharyya y Wilcoxon. A partir de estos rankings se han seleccionado únicamente las características que mejoraban la probabilidad de acierto final.

También es de destacar que se han realizado 3 experimentos de clasificación. Los 2 primeros se corresponden con cada uno de los criterios de detección de  $L_T$ , estos son, el volumen y la superficie. De ambos criterios surgen valores distintos de  $L_T$ , y por tanto, distintas matrices de características de clasificación. El tercero de ellos surge al unir las matrices de parámetros de los 2 experimentos anteriores.

## VI. RESULTADOS

En primer lugar, podemos visualizar cual es la distribución de las isosuperficies umbrales para las imágenes HC y PD en función del criterio seleccionado, esto es, volumen o superficie.

Sin embargo, como las imágenes no presentan la misma intensidad máxima, para dar una mejor medida comparativa, en la Figura 9 se muestran los valores de  $L_T$  normalizados por la máxima intensidad de cada imagen, agrupados en un *boxplot*. La línea roja marca la mediana de la distribución de intensidades, los límites azules los percentiles 25 y 75, y los límites discontinuos los valores máximos y mínimos no

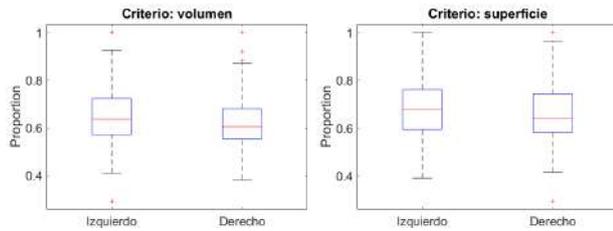


Fig. 9: Boxplots de las isosuperficies umbrales de cada imagen normalizadas.

atípicos. Por el contrario, los puntos más allá de estos límites son los valores atípicos.

Por otra parte, dado que hemos empleado 2 algoritmos de clasificación, en las Tablas IV y V se muestran los mejores resultados de clasificación en función del experimento de clasificación llevado a cabo. Dichos resultados se han calculado como el promedio de 100 realizaciones en el caso de SVM y de 5 realizaciones para Adaptive Boosting. Los resultados se muestran de la forma  $\bar{m} \pm \sigma$ .

	Accuracy	Sensitivity	Specificity	Precision
<b>Surface</b>	92.4 $\pm$ 0.8%	92.9 $\pm$ 1.3%	91.9 $\pm$ 1.1%	91.4 $\pm$ 1.1%
<b>Volume</b>	91.5 $\pm$ 1.4%	94.9 $\pm$ 1.3%	88.3 $\pm$ 2.0%	88.3 $\pm$ 1.8%
<b>Both</b>	91.9 $\pm$ 1.3%	95.0 $\pm$ 1.4%	89.0 $\pm$ 1.9%	88.9 $\pm$ 1.7%

Tabla IV: Resultados de clasificación empleando como clasificador una SVM.

	Accuracy	Sensitivity	Specificity	Precision
<b>Surface</b>	91.9 $\pm$ 0.7%	90.4 $\pm$ 1.8%	93.3 $\pm$ 2.1%	92.7 $\pm$ 2.0%
<b>Volume</b>	91.8 $\pm$ 0.7%	88.6 $\pm$ 1.7%	94.8 $\pm$ 0.8%	94.1 $\pm$ 0.8%
<b>Both</b>	92.0 $\pm$ 2.1%	91.2 $\pm$ 3.8%	92.8 $\pm$ 2.1%	92.2 $\pm$ 2.2%

Tabla V: Resultados de clasificación empleando como clasificador el algoritmo de Adaptive Boosting.

Finalmente, podemos representar estos resultados en una curva ROC, tal y como muestra la Figura 10, donde también se indica el área bajo la curva de cada gráfica.

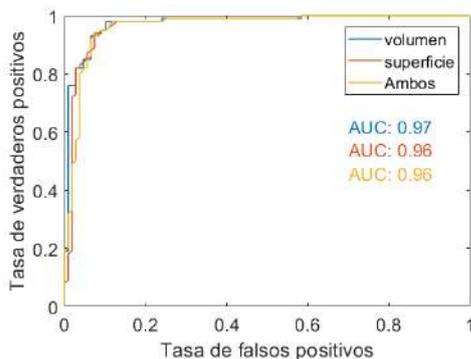


Fig. 10: Curvas ROC para los mejores resultados de cada criterio.

## VII. DISCUSIÓN

En primer lugar, si observamos los boxplots de la Figura 9, la distribución de  $L_T$  normalizada no presenta diferencias significativas para ambos criterios, lo que justifica que en las Tablas IV y V tampoco haya grandes cambios en los

resultados de clasificación entre ambos criterios. Asimismo, si comparamos los 2 algoritmos de clasificación empleados, vemos que en general los resultados son similares, donde el mejor resultado de precisión se ha obtenido mediante el criterio de la superficie y empleando una SVM.

En comparación con otros estudios, como [3], los resultados de clasificación son inferiores, debido a que en este trabajo se emplean características halladas a partir de una única isosuperficie, mientras que en trabajos previos se concatenaban las características de muchas isosuperficies. No obstante, esta reducción es compensada con una disminución del número de características (se evita overfitting) y un menor coste computacional.

## VIII. CONCLUSIONES

A partir del desarrollo realizado y los resultados obtenidos, podemos afirmar que los objetivos enunciados al principio del documento han sido cumplidos, dado que se han obtenido unos resultados de clasificación comparables a los de la comunidad científica sin aplicar ningún tipo de normalización espacial a las imágenes. Además, se ha desarrollado un nuevo método de obtención de una isosuperficie umbral que contenga la información más relevante del cuerpo estriado para clasificar a los pacientes. Asimismo, se han extraído parámetros a partir de dicha isosuperficie que han permitido caracterizar al cuerpo estriado de cada sujeto.

Estos resultados podrían suponer un nuevo paradigma en el procesamiento de imágenes funcionales, puesto que este trabajo es uno de los primeros que obtiene buenos resultados sin aplicar una normalización espacial a las imágenes, manteniendo inalteradas las características del cuerpo estriado. Con lo cual, las diferencias entre pacientes se deben realmente a diferencias en la cantidad de células dopaminérgicas, permitiendo aunar la interpretación médica con la computacional.

## IX. FUTURAS LÍNEAS DE INVESTIGACIÓN

A pesar de los buenos resultados obtenidos en este trabajo, la idea propuesta podría ampliarse en un futuro próximo. Teniendo en cuenta la labor de investigación realizada, las posibles mejoras futuras de este trabajo son las siguientes:

- 1) Planteamiento de otros métodos de cálculo de  $L_T$ , de modo que se puedan comparar los resultados con los presentados aquí.
- 2) Extracción de más características morfológicas del estriado que permitan mejorar la tasa de acierto del clasificador.
- 3) Análisis de las características extraídas mediante otros clasificadores o redes neuronales.

## AGRADECIMIENTOS

A mis padres y hermanos, por su apoyo durante todos mis años de estudio, y a mis profesores, en especial a mi tutor, por su esfuerzo en transmitir los conocimientos que nos hacen profesionales.

## REFERENCIAS

- [1] D. Castillo-Barnes, et al. "Comparison Between Affine and Non-affine Transformations Applied to  $I^{[123]}$ -FP-CIT SPECT Images Used for Parkinson's Disease Diagnosis", en *Understanding the Brain Function and Emotions*, Cham, 2019, págs. 379-388.
- [2] Francisco M.P. Oliveira et al. (2015, febrero). "Computer-aided diagnosis of Parkinson's disease based on  $[123I]$ FP-CIT SPECT binding potential images, using the voxels-as-features approach and support vector machines", *Journal of Neural Engineering*. 12(2)
- [3] D. Castillo-Barnes, et al., "Analysis of  $I[123]$ -Ioflupane SPECT intensity iso-surfaces to assist the diagnosis of Parkinsonism", 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings NSS/MIC, 2018.
- [4] Junta de Andalucía, "Retribuciones del personal de Centros e Instituciones Sanitarias", 2020. Disponible en: [https://www.sspa.juntadeandalucia.es/servicioandaluzdesalud/sites/default/files/sincfiles/wsas-media-sas\\\_normativa\\\_mediafile/2020/11881.pdf](https://www.sspa.juntadeandalucia.es/servicioandaluzdesalud/sites/default/files/sincfiles/wsas-media-sas\_normativa\_mediafile/2020/11881.pdf)
- [5] Pablo Padilla et al. (2015) "Intensity normalization in the analysis of functional datscan spect images: The  $\alpha$ -stable distribution-based normalization method vs other approaches". *Neurocomputing*. 150, págs. 4-15
- [6] Edelsbrunner, Herbert and Kirkpatrick, David and Seidel, Raimund (1983). "On the shape of a set of points in the plane". *IEEE Transactions on information theory*. 29(4), págs. 551-559.

# Uso de líneas de contorno en imágenes [123I] FP-CIT SPECT para el diagnóstico de Parkinson

Autor: Mario Molina Molina, e-mail: mariomolinamolina1994@gmail.com

Tutores: Diego Salas González (dsalas@ugr.es), Javier Ramírez Pérez de Inestrosa (javierrp@ugr.es), Juan Manuel Górriz Sáez (gorriz@ugr.es)

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—El Parkinson es una enfermedad neurodegenerativa que daña el sistema motor de forma crónica y progresiva. Suele afectar a personas de avanzada edad y ya se ha convertido en una de las enfermedades más comunes, en parte por el envejecimiento de la población. En su diagnóstico se observa que causa la muerte de células del cerebro en algunas áreas, en las que se detecta la pérdida de dopamina.

En este trabajo se va a presentar el procesado de imágenes biomédicas cerebrales de tipo I-123 SPECT para el diagnóstico del Parkinson. El [123I] FP-CIT SPECT (DATSCAN) es un radiotrazador que se une a los transportadores de dopamina en el área estriatal. Según estudios, los niveles de baja intensidad en el cuerpo estriado se relacionan con la enfermedad.

Se van a extraer características relacionadas con la geometría del cuerpo estriado, en las dos proyecciones obtenidas a partir de las imágenes proporcionadas en 3D por el Hospital Virgen de las Nieves de Granada. Estas imágenes están etiquetadas por expertos como pacientes sanos (NOR) y pacientes con Parkinson (PKS). A partir de estas características, combinadas con alguna otra relacionada con la intensidad, se va a entrenar y clasificar mediante Máquina de Vectores de Soporte (SVM).

Se demostrará que el uso de características relacionadas con la morfometría del estriado, pueden mejorar los resultados de clasificación obtenidos mediante métodos que sólo tienen en cuenta la intensidad de los píxeles o vóxeles del cuerpo estriado en imágenes neuro-cerebrales.

**Palabras clave**—contornos, Parkinson, imágenes biomédicas, clasificación, SVM.

## I. INTRODUCCIÓN

EL Parkinson (conocido por las iniciales EP o PD, en inglés) es una enfermedad neurodegenerativa que afecta al sistema motor, y que suele afectar a personas de avanzada edad, convirtiéndose en una de las enfermedades más comunes hoy en día. Esta enfermedad causa la muerte de células del cerebro en algunas áreas, en las que se detecta la pérdida de dopamina.

Se va a presentar el procesado de imágenes biomédicas cerebrales de tipo I-123 SPECT para el diagnóstico del Parkinson. El [123I] FP-CIT SPECT (DATSCAN) es un radiotrazador que se une a los transportadores de dopamina en el área estriatal. Por lo tanto, los niveles de baja intensidad en el cuerpo estriado se relacionan con la EP. [1], [2]. En general, las imágenes de pacientes con Parkinson en la base de datos tiene menos intensidad que las de pacientes sanos, como se puede observar en las figuras 1 y 2.

Este sistema para la clasificación de Parkinson surge por la necesidad de realizar un examen de detección de esta

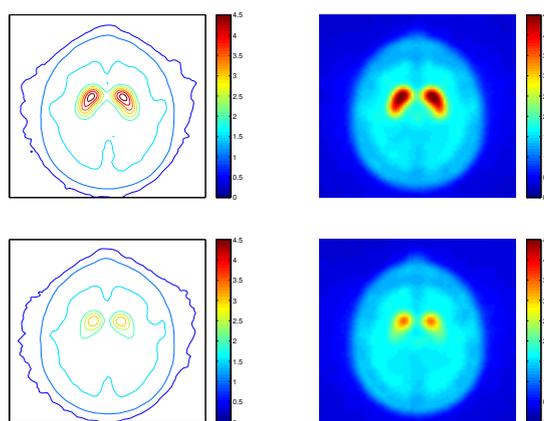


Fig. 1. Contornos (izquierda) e intensidad (derecha) de la media de las imágenes en corte axial NOR (arriba) y PKS (abajo)

enfermedad cuando aún no existen síntomas físicos evidentes o el profesional médico tiene dudas acerca del diagnóstico.

En muchos casos, la enfermedad de Parkinson se puede confundir con otras enfermedades similares o temblores propios de la avanzada edad, por lo que el análisis visual de las imágenes cerebrales puede no ser suficiente.

## II. MATERIALES Y MÉTODOS

### A. Base de datos

Se obtuvieron 147 imágenes tras un período de 3-4 horas de la inyección intravenosa de 185 MBq (5 mCi) de 123 I-FP CIT, y posteriormente, bloquear la tiroides con la solución de Lugol. La dimensión de la imagen resultante fue de 128 x 128 x 60 y la dimensión del píxel de 3.2 mm. Las imágenes fueron etiquetadas por un comité de expertos del Hospital Virgen de las Nieves de Granada, España. El resultado fueron: 88 pacientes con controles sanos normales (etiquetados como NOR) y 59 pacientes con la enfermedad de Parkinson (etiquetados como PKS).

Con el objetivo de ampliar la zona de interés, se realizó una normalización espacial en la que las imágenes resultantes tienen una dimensión de 73 x 73 x 45, mucho más manejable y compacta para realizar los cálculos necesarios en la extracción de las características. De este modo, tras la normalización espacial, las mismas coordenadas en la imagen normalizada

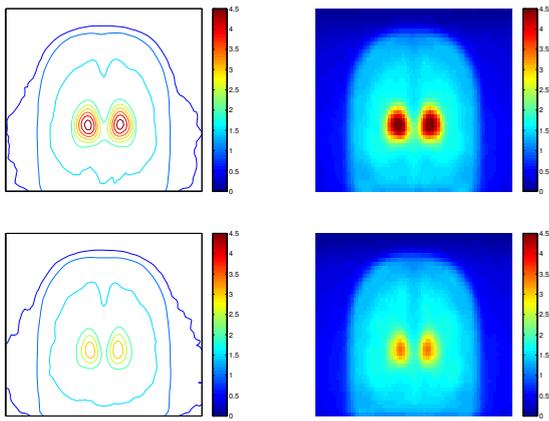


Fig. 2. Contornos (izquierda) e intensidad (derecha) de la media de las imágenes en corte coronal NOR (arriba) y PKS (abajo)

se corresponde con las mismas posiciones espaciales en el cerebro.

Las imágenes se encuentran en formato *Analyze* y se implementa una función en Matlab para leerlas y transformarlas en una matriz de 4 dimensiones. La primera dimensión se refiere al número de imágenes de la base de datos (147 en nuestro caso), mientras que el resto de dimensiones se refiere a la propia dimensión de la imagen cerebral en 3D (73 x 73 x 45). Por tanto, la matriz resultante tendrá unas dimensiones de 147 x 73 x 73 x 45.

### B. Normalización en intensidad de las imágenes

La intensidad de las imágenes se ha normalizado según la relación entre el valor de intensidad en cada vóxel y el promedio de la intensidad en la región no específica (región occipital).

En la figura 3 se puede observar la dimensión de la máscara, para los distintos cortes o proyecciones, que ocupa esta región occipital, donde no existe información específica sobre la enfermedad y, por lo tanto, podemos considerar que los valores de intensidad obtenidos allí son debidos a ruido aleatorio. La máscara se trata de un prisma ubicado en esta región. En el apartado de descripción del modelo, se detallarán las funciones y procedimientos usados.

Por lo tanto, se calcula la relación de unión específica (más conocido como Binding Ratio, BR). Para calcular el promedio de la intensidad en esa región, se ha calculado una máscara en 3D visualmente con la ayuda de la herramienta de Matlab *SliceBrowser*.

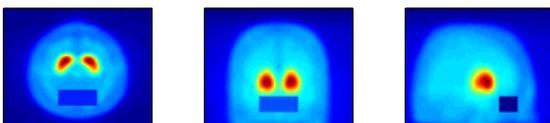


Fig. 3. Máscara usada en la normalización en intensidad de la base de datos de las imágenes 3D, correspondiente a la región occipital (no específica). Vista desde el corte axial (primera), coronal (segunda) y sagital (tercera)

### C. Proyección axial y coronal

En primer lugar, antes de la extracción de características, a partir de cada una de las 147 imágenes en 3D de la base de datos, se proyecta el máximo en cada corte transaxial en una imagen cerebral 2D. En segundo lugar, también proyectamos el máximo en el corte coronal. Por lo tanto, la base de datos en estudio consistirá en 2 x 147 imágenes 2D, divididas en corte axial y coronal. El corte sagital se va a descartar del estudio y de la extracción de características, puesto que se tendrá menos información del cuerpo estriado en este corte de la que tienen los demás. Esto es debido a que sólo se puede observar una parte del cuerpo estriado.

Las figuras 1 y 2 representan las líneas de contorno y los valores de intensidad de la media de las imágenes en corte axial y coronal de pacientes con Parkinson y de pacientes sanos. Se puede observar la región específica como el área estriada (o de mayor intensidad). Esta figura también muestra las principales diferencias entre las imágenes promedio PKS y NOR:

- Intensidad (o nivel de contorno) del cuerpo estriado.
- Forma geométrica del cuerpo estriado.

El valor de intensidad del cuerpo estriado en pacientes con Parkinson es menor que el valor de intensidad en pacientes sanos. Este resultado también era esperado, debido a que, tal y como apuntábamos en la introducción, los niveles de dopamina medidos en pacientes con Parkinson es menor que en pacientes sanos. Por otro lado, de igual manera, la forma geométrica del cuerpo estriado es diferente en pacientes con Parkinson: el área que delinea el cuerpo estriado es menor y más redondeada en pacientes con Parkinson que en pacientes sanos, con una forma más ovalada (sobre todo en la proyección axial). Esto es debido también a que la absorción de dopamina es menor en pacientes con Parkinson que en pacientes sanos.

### D. Niveles de contorno

Los líneas de contorno son un conjunto de valores con el mismo nivel o intensidad y que pueden constituir una forma geométrica. Para este caso, vamos a usar aquellas líneas de contorno que forman el cuerpo estriado, y por tanto tienen mayor nivel de contorno. Los valores de intensidad o niveles de contorno son más bajos en el estriado de las imágenes de pacientes con Parkinson. También, la forma del área estriada es más redondeada.

Empleando la información proporcionada por la proyección axial y coronal, estimamos el valor de intensidad que delinea el área del estriado izquierdo y derecho. Estos valores serán nuestra referencia de intensidad para una imagen dada. Una vez que se ha calculado este valor, elegimos valores diferentes de intensidad en aumento e igual espaciados, hasta un valor máximo en el área del cuerpo estriado izquierdo y derecho en la imagen. Este valor máximo será el 95% del valor de intensidad máximo de cada imagen de la base de datos, para que se consiga un contorno máximo que pueda ocupar más de un píxel.

Por lo tanto, finalmente obtenemos 4 líneas de contorno en cada parte del cuerpo estriado. En la figura 4 se muestran un ejemplo de las 4 líneas de contorno en cada cuerpo estriado medida para una imagen PKS y otra NOR.

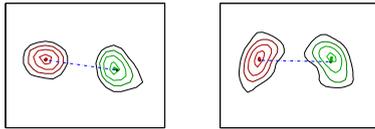


Fig. 4. Ejemplo de líneas de contornos de imágenes PKS (arriba) y NOR (abajo) para proyección axial. Los asteriscos en el interior de los contornos representan los centros de estos, mientras que la línea discontinua azul representa la distancia entre centros. A cada parte del estriado se le ha asignado un tono diferente para diferentes valores de nivel.

### E. Extracción de características

Para la extracción de características, vamos a dividir nuestras imágenes en dos mitades, de forma que el cuerpo estriado quede en dos partes (izquierda y derecha). Una vez obtenidas las diferentes características para las dos partes, se le aplica una media y se calculan características de relación entre las dos partes.

Realizamos la extracción de características utilizando la información proporcionada por las 4 líneas de contorno calculadas en el cuerpo estriado para cada parte y para las dos bases de datos de las imágenes en 2D consideradas: las proyecciones axial y coronal. Las características calculadas en cada parte del estriado, para cada una de las líneas de contorno son:

- 1) Área (A)
- 2) Perímetro (P)
- 3) Redondez (R): medida como el área dividida por el cuadrado del perímetro,

$$R = \frac{A}{P^2} \quad (1)$$

- 4) Radio de las líneas de contorno (r): como el radio de la circunferencia circunscrita a la línea de contorno. Este radio, por tanto será la distancia entre el centro y el punto del contorno más alejado.
- 5) Nivel de líneas de contorno: relacionado con la intensidad en esa línea de contorno.
- 6) Posición de los centros de las líneas de contorno ( $x_c$ ,  $y_c$ ).
- 7) Número de líneas de contorno fuera y dentro del estriado: nuestro objetivo es seleccionar solo aquellas 4 líneas de contorno que describe el cuerpo estriado para los cálculos geométricos anteriores. A pesar de ello, se podrán formar puntos o contornos fuera del estriado en algunas imágenes y que sólo nos interesaría saber el número de ellas. Esto es porque en algunos casos, a mayor número de líneas de contorno, mayor posibilidad de diagnóstico positivo en la clasificación. Se estudiará en los siguientes apartados.

A partir de ellas, calculamos la media, mínimo, máximo o diferencias de todas estas características en cada parte del estriado (izquierda y derecha). Después, vamos a obtener cada característica para la imagen general, haciendo la media de todas las características anteriores de la parte derecha e izquierda. Además, también se mide la distancia entre cada contorno de la parte derecha e izquierda, relación de áreas o alineación de los centros, así como la media entre todos estos contornos. De esta forma, las características generales que se calculan son:

- 1) Distancia entre las partes del contorno (D): se calcula como la distancia entre los centros de cada línea de contorno.

$$D = \sqrt{(x_{c_{izq.}} - x_{c_{der.}})^2 + (y_{c_{izq.}} - y_{c_{der.}})^2} \quad (2)$$

- 2) Relación de áreas ( $R_A$ ): consiste en la relación entre las áreas de las líneas de contorno de cada parte del estriado. Este valor variará entre 0 y 1.
- 3) Alineación de centros: para cada coordenada, X e Y, se realiza una relación entre las coordenadas los centros de los contornos del estriado derecho e izquierdo.

En total se calculan 68 características (entre medias, máximas o mínimas) de los contornos de la parte derecha e izquierda como las áreas, redondez, radios y centros de los contornos, distancia entre los contornos de ambas partes o niveles de los contornos. Se buscan las características más discriminatorias, esto es, aquellas que puedan diferenciar mejor una imagen cerebral de un paciente con Parkinson y otra de uno sano.

Una de estas características discriminatorias en Parkinson, es la diferencia entre las dos partes del cuerpo estriado (o asimetría), tanto en intensidad como en geometría. Es decir, las imágenes de pacientes con Parkinson, suelen tener menor intensidad en la media de los contornos, así como mayor diferencia entre la parte derecha e izquierda del estriado. Lo mismo ocurre en aspectos geométricos: las líneas de contorno en imágenes de pacientes con Parkinson suelen tener una mayor forma redondeada y de área menor que en las de pacientes sanos. Esto se puede observar en las imágenes de las líneas de contorno de la figura 4.

### F. Clasificación SVM

Realizamos la tarea de clasificación utilizando Support Vector Machine (SVM) con kernel lineal, cuadrático, polinomial y RBF (Función de Base Radial Gaussiana). Además, para calcular el rendimiento de la clasificación se obtiene el acierto utilizando una estrategia de validación cruzada *Leave-One-Out*. Vamos a comparar estos resultados de dos formas: en términos de porcentaje de acierto o en términos de coste computacional.

Para el primero de ellos, calculamos el porcentaje de acierto, especificidad y sensibilidad cuando solo se utiliza una de las características. De esta forma, podemos ordenar cada característica de acuerdo con su rendimiento. Más tarde, combinamos las características en orden creciente, según su porcentaje de acierto: la primera; la primera y la segunda; la primera, la segunda y la tercera; y así sucesivamente. Y calculamos, una vez más, el rendimiento de la clasificación cuando se combinan las características más significativas como entrada en SVM. Por otro lado, se van a ordenar también las características mediante *rankfeatures*.

Para el segundo método, se seleccionan sólo características medias del estriado catalogadas como más significativas o discriminatorias, reduciendo de este modo la dimensión de las características usadas.

## III. RESULTADOS Y DISCUSIÓN

De los resultados (tabla I) podemos deducir que la proyección axial se adapta mejor a la función kernel RBF, mientras que la proyección coronal a la función lineal.

Tabla I  
RESULTADOS DE ACIERTO EN LOS DIFERENTES KERNELS DE SVM PARA EL CORTE AXIAL Y CORONAL

	Lineal	Cuadrático	Polinomial	RBF
Axial	0.8571	0.8367	0.5986	0.8912
Coronal	0.8231	0.7755	0.5986	0.8095

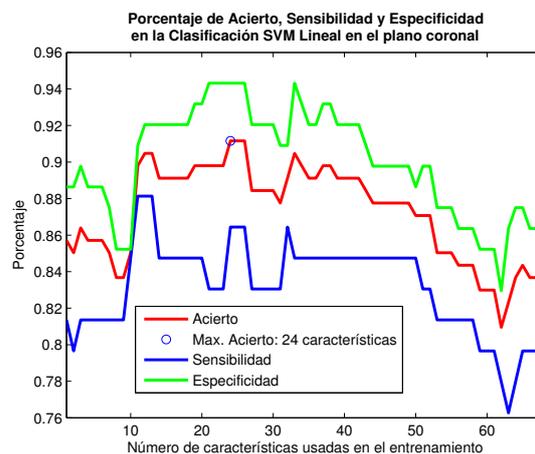
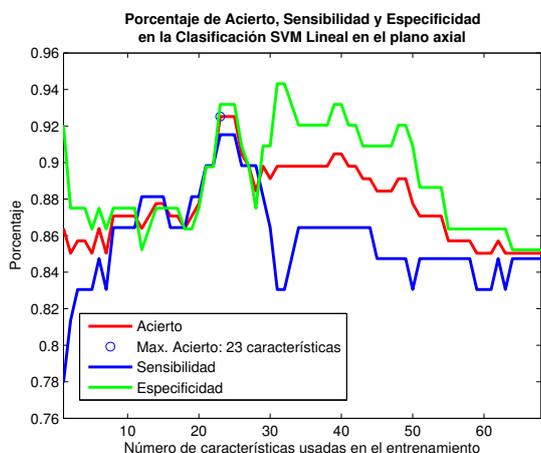


Fig. 5. Porcentaje de acierto, sensibilidad y especificidad de la Clasificación SVM lineal en plano axial

Fig. 7. Porcentaje de acierto, sensibilidad y especificidad de la Clasificación SVM lineal en plano coronal

Con el objetivo de optimizar los resultados, evaluamos el porcentaje de acierto de cada característica individualmente. Posteriormente, ordenamos las características en orden descendente según este porcentaje de acierto y elegimos cuales de ellas son las más significativas. Para la elección de las características más significativas, se han calculado diferentes resultados, para los dos kernels que mejores resultados daban en la proyección coronal y axial (kernel de función lineal y RBF). Conforme se aumentaba el número de características según su porcentaje de acierto individual, se obtenía un nuevo porcentaje de acierto para las características seleccionadas (observadas en las figuras 5, 6, 7 y 8). A partir de las gráficas, se obtienen los valores óptimos de porcentaje de acierto para un número de características concreto.

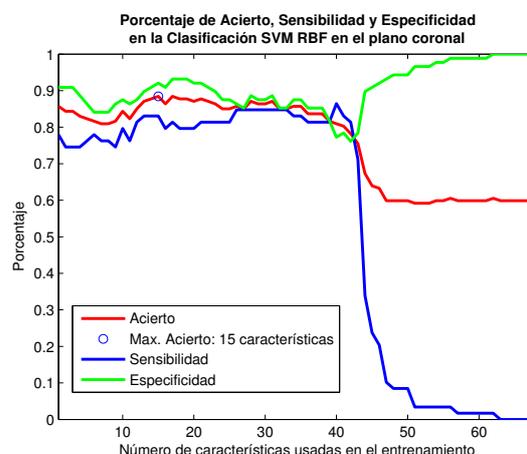


Fig. 8. Porcentaje de acierto, sensibilidad y especificidad de la Clasificación SVM RBF en plano coronal

En el caso de la proyección axial, el mejor resultado para

el porcentaje de acierto con el kernel lineal es de 0.9252 usando las 23 características más significativas (relacionadas con el radio, el nivel, el área y la redondez de los contornos, respectivamente) y con kernel RBF de 0.8980 usando 5 características (relacionadas con el radio, el área y el nivel). Para la proyección coronal, el mejor resultado del porcentaje de acierto es de 0.9116 usando 24 características con el kernel lineal y con el kernel RBF 0.8844 usando 15 características (relacionadas con el nivel, el área y la coordenada x de los centros de los contornos, para ambos casos).

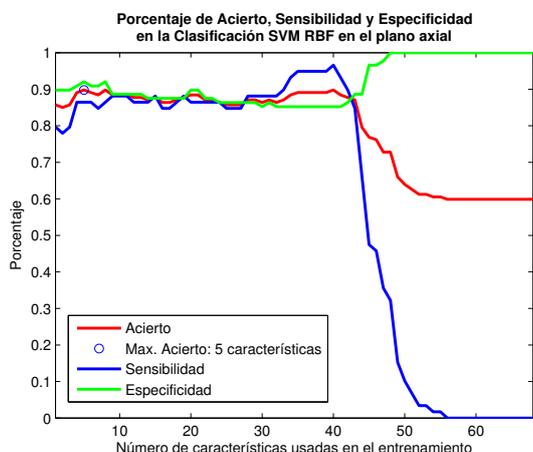


Fig. 6. Porcentaje de acierto, sensibilidad y especificidad de la Clasificación SVM RBF en plano axial

IV. CONCLUSIÓN

En este trabajo, se propone un método de extracción de características para la implementación de un sistema de diagnóstico asistido por computadora para realizar la clasificación de una base de datos con pacientes con controles sanos

	Axial	Coronal
Lineal	0.9252	0.8980
RBF	0.9116	0.8844

Tabla II

PORCENTAJE MÁXIMO DE ACIERTO PARA LA PROYECCIÓN AXIAL Y CORONAL USANDO EL KERNEL LINEAL Y RBF

y pacientes con EP. Extraemos características de las imágenes [123I] FP-CIT SPECT usando curvas de nivel. Por lo tanto, la forma de las regiones cerebrales funcionales no está limitada a un vóxel.

La metodología propuesta se prueba en un conjunto de 147 imágenes cerebrales en 3D, de las que se obtienen dos nuevas bases de datos (en 2D), por la proyección transaxial y coronal de los píxeles con mayor intensidad. Los resultados de rendimiento muestran que la inclusión de características basadas en formas mejora la tarea de clasificación en comparación con el caso en el que solo se utilizan las características de valores de intensidad. Frente al 89.12% de acierto obtenido en el método de intensidad, se obtiene 92.52% y 93.20% en los métodos propuestos. En la tabla III se resumen los mejores resultados obtenidos con todos los métodos, en los que se pueden comparar diferentes parámetros de la clasificación.

Método		Acierto	Caract.	Corte	Kernel
Intensidad		0.8912	1	Axial	Polynomial
Completo		0.8912	68	Axial	RBF
Opt.	Ordenación	0.9252	23	Axial	Lineal
	Rankfeatures	<b>0.9320</b>	33	Axial	Lineal
Simp.	Ordenación	<b>0.9320</b>	5	Axial	RBF
	Rankfeatures	<b>0.9320</b>	3	Axial	Polynomial

Tabla III

MEJORES RESULTADOS DE TODOS LOS MÉTODOS USADOS

Se obtiene el mejor resultado de acierto (de 93.20%) para varios métodos usados, en los que varía el número de características óptimas usadas en la clasificación, el corte o proyección y el kernel. En general, se puede observar que usando la proyección axial se obtienen los resultados óptimos.

También se ha demostrado que no por el uso de más características en el entrenamiento, el acierto será mayor, si no que basta con el uso de características discriminatorias basadas en geometría conjuntamente con las relacionadas con los niveles de intensidad. De hecho, se obtiene el mismo resultado cuando usamos las 8 características más significativas y reducimos las dimensiones de las características en el método simplificado.

#### AGRADECIMIENTOS

En primer lugar, agradecer a mis padres todo el apoyo que me han dado durante todos estos años, sin ellos no hubiera sido posible nada de esto. Al resto de mi familia y amigos, que siempre han estado para darme ánimos, y como no, a mis tutores, por permitirme y ayudarme a realizar este trabajo.

#### REFERENCIAS

- [1] Salas-Gonzalez D., Gorriz J. M., Ramirez J., I. Illan y E. Lang, "E.Linear intensity normalization of FP-CIT SPECT brainimages using the alpha-stable distribution". *NeuroImage* 65(2013), 449–455.
- [2] Hauser, R. A. y Grosset, D.G. "[123I]FP-CIT (DaTscan)SPECT brain imaging in patients with suspected parkinsonian syn-dromes". *Journal of Neuroimaging* 22, 3 (2011), 225–230.



# USO DE WAVELETS PARA REDUCCIÓN DE RUIDO: APLICACIÓN EN SEÑALES DE VOZ Y SEÑALES SÍSMICAS

Autora: Laura Gámiz Pérez, e-mail: lauragamiz@correo.ugr.es

Tutora: M<sup>a</sup> Carmen Benítez Ortúzar; e-mail: carmen@ugr.es

Tutora: Luz García Martínez, e-mail: luzgm@ugr.es

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones

Universidad de Granada

**Resumen**—En este Proyecto se analizan e implementan técnicas de reducción de ruido para señales reales unidimensionales, en concreto se centra en señales de voz y en señales de naturaleza volcano-tectónica. Estas técnicas hacen uso de la Transformada Wavelet Discreta, a partir de la cual se realiza una descomposición en niveles de la señal. Para ello se utilizan filtros paso-alto y paso-baja, que son los responsables de obtener los coeficientes sobre los que se trabaja. Lo que se pretende es eliminar aquellos coeficientes que se corresponden con el ruido de fondo mientras se mantienen aquellos que hacen referencia a parte significativa de la señal. Además, se ha desarrollado una nueva técnica que se adapta a las características de las señales objeto de estudio y que acaba con algunos de los problemas que presentan las técnicas ya existentes. Para comprobar el funcionamiento de las técnicas se ha realizado un estudio comparativo de las señales obtenidas tras la aplicación de la mismas y de las señales ruidosas. Además, para la valoración de cada algoritmo se ha realizado un test MOS en el caso de señales de voz y se ha usado un algoritmo automático de detección del *picking* para señales sísmicas.

**Palabras clave** — Transformada Wavelet, señales de voz, señales sísmicas, reducción de ruido.

## I. INTRODUCCIÓN

EN el campo de la información y las telecomunicaciones es de gran importancia el estudio de señales de distinta naturaleza, pues éstas reflejan el comportamiento de los fenómenos físicos implicados. Sin embargo, en cualquier sistema de comunicación, la señal que se recibe difiere de la señal transmitida debido a diversos factores. Entre ellos, el ruido es uno de los más notorios y perjudiciales. Pero, eliminar el ruido presente en señales no es una tarea fácil, pues existe una amplia variedad de fuentes que lo provocan y puede ser el resultado de la combinación de varios factores. Además, el ruido puede ser intermitente o afectar únicamente a determinadas componentes frecuenciales, lo que dificulta su estudio y su tratamiento.

La reducción de ruido en señales de voz ha sido objeto de numerosas investigaciones a lo largo de la historia. Asimismo, sigue siendo de gran interés en la actualidad debido al elevado número de aplicaciones de las que forman parte este tipo de señales: comunicaciones móviles, reconocimiento automático del habla, domótica, etc.

Por otro lado, la sismología es la ciencia que se encarga del estudio de la estructura interna de la Tierra, así como del origen y propagación de terremotos. En los últimos años, la predicción de terremotos originados por la actividad volcánica

y la determinación de futuros procesos eruptivos u otros riesgos geológicos ha sido una de sus principales líneas de investigación. Así pues, la detección de los tiempos de llegada de las distintas ondas sísmicas (*picking*) registradas en los sismogramas da información sobre la topología y el interior de la tierra. Además, en el caso de volcanes, los cambios en la velocidad de las ondas asociadas están estrechamente relacionados con la densidad del terreno, por lo que dicha detección puede informarnos de la existencia de magma o anomalías. Para que los datos analizados sean fiables y efectivos es imprescindible disponer de señales limpias y sin distorsiones.

Con este propósito se pretende encontrar nuevas técnicas que reduzcan el ruido en ambos tipos de señales y que acaben con las deficiencias de los métodos de reducción de ruido tradicionales. Así pues, el principal objetivo es la implementación en MATLAB de técnicas de reducción de ruido basadas en el procesado y análisis Wavelets de señales, lo que se conoce como *Wavelet Denoising* o *Wavelet thresholding*. Estas técnicas serán aplicadas a señales de voz y a señales sísmicas y se realizará un estudio del comportamiento de las mismas en ambos tipos de señales.

## II. ESTADO DEL ARTE

### A. La señal de voz

La voz es una onda acústica de presión generada por movimiento fisiológicos voluntarios de los órganos del aparato fonador humano.

Se trata de una señal de banda ancha con frecuencias que varían desde los 100 Hz a los 8000 Hz aproximadamente. No obstante, la mayor parte de la energía está concentrada por debajo de los 4 kHz, por lo que la inteligibilidad de la voz tiene mayor peso en este rango de frecuencias. Además, es una señal no estacionaria, pues sus componentes en tiempo y en frecuencia se encuentran en continuo cambio. Asimismo, esta señal presenta una estructura compleja, lo que dificulta su estudio y tratamiento.

Existen distintos factores que pueden afectar a la señal de voz, impidiendo su correcta inteligibilidad y la extracción de determinadas características de la misma. Entre estos efectos nocivos se encuentra el ruido, el cual es totalmente independiente de la señal y puede producir una degeneración de la misma.

Para el proceso de reducción de ruido es fundamental conocer las características de la señal afectada, así como las

del ruido y las fuentes que lo producen. Así pues, algunas de las fuentes predominantes en la señal de voz, y con las que hemos trabajado en este proyecto son el ruido del tráfico y el ruido procedente de voces de otras personas, también conocido como *babble*.

**B. Señales sismo-volcánicas**

Los sismos volcano-tectónicos o VT son terremotos que ocurren en un ambiente volcánico. Las señales que definen a estos eventos se caracterizan por presentar un contenido frecuencial que abarca de 1 a 40 Hz. Además, son señales con duración y magnitud variables.

Las señales sismo-volcánicas están formadas por dos tipos de ondas: ondas de cuerpo y ondas de superficie. Las ondas de cuerpo se transmiten en el interior de la Tierra y emergen hacia la superficie. Estas ondas se dividen, a su vez, en ondas P y S. Por el contrario, las ondas superficiales se propagan a lo largo de la superficie libre de la Tierra o en zonas de discontinuidad del interior de la misma. Estas últimas se dividen en ondas Rayleigh y ondas Love.

En una medición sísmica existen tres componentes de movimiento, que son detectadas por los sismómetros y registradas en los sismogramas. Así pues, cada terremoto VT tendrá asociadas tres trazas, correspondientes a las componentes Norte-Sur, Este-Oeste y Vertical. En algunas de estas componentes se podrán detectar con mayor claridad los distintos tipos de ondas mencionados. En todos los terremotos volcano-tectónicos la primera en llegar es la onda P, seguida de la onda S y de las ondas superficiales. La diferencia en los tiempos de llegada de las ondas P y S se puede utilizar para determinar las características del terreno que atraviesan y la existencia de anomalías en el mismo. No obstante, cuando el ruido presente en la señal es muy elevado, la distinción de estas ondas se vuelve una tarea ardua y tediosa.

Existen un amplio rango de factores que pueden contribuir al ruido de fondo en señales sísmicas. Entre estos se incluyen la presencia de mares, presión atmosférica, efectos asociados a la variación de la temperatura, el viento o la actividad inducida por el hombre. Clasificar todas estas fuentes de ruido no es una tarea fácil. Por este motivo, numerosos autores han utilizado el rango de frecuencia al que afectan las distintas fuentes como criterio de clasificación [1].

**C. Transformada Wavelet Discreta**

La Transformada Wavelet fue desarrollada con el objetivo de acabar con las limitaciones que presentaba la Transformada de Fourier. Al igual que la Transformada de Fourier de Tiempo Corto, permite realizar un análisis simultáneo tiempo-frecuencia pero, a diferencia de esta última, usa ventanas con un ancho variable.

El análisis wavelet se basa en la superposición de una señal en versiones desplazadas y escaladas de la wavelet madre. Así pues, la Transformada Wavelet Discreta (DWT) de una señal discreta,  $f[n]$ , queda definida de la siguiente forma:

$$c_{j,k}[n] = \frac{1}{\sqrt{2^j}} \sum_k f[n] \psi\left(\frac{n-k2^j}{2^j}\right) \quad (1)$$

donde  $\psi_{j,k}$  es una función wavelet que viene dada por la expresión (2).

$$\psi_{j,k}[n] = \frac{1}{\sqrt{2^j}} \psi\left(\frac{n - k2^j}{2^j}\right) \quad (2)$$

La discretización abordada por la Transformada Wavelet Discreta (DWT) permite representar una señal como funciones elementales acompañadas de coeficientes. Esto da lugar, a dos tipos de funciones base, Wavelets Madre y funciones de escala, mediante las cuales se pueden obtener los coeficientes que caracterizan a la señal. Las Wavelets Madre representan los detalles finos de la señal, mientras que las funciones de escala son las encargadas de realizar una aproximación de la misma. Así pues, podemos aproximar una función discreta como una sumatoria de funciones wavelet y funciones de escala, tal y como aparece en la expresión (3).

$$f[n] = \frac{1}{\sqrt{M}} \sum_k c_\phi[j_0, k] \phi_{j_0, k}[n] + \frac{1}{\sqrt{M}} \sum_k d_\psi[j, k] \psi_{j, k}[n] \quad (3)$$

siendo los coeficientes de aproximación y de detalle, respectivamente,

$$c_\phi[j_0, k] = \frac{1}{\sqrt{M}} \sum_n f[n] \phi_{j_0, k}[n] \quad (4)$$

$$d_\psi[j, k] = \frac{1}{\sqrt{M}} \sum_n f[n] \psi_{j, k}[n] \quad j \geq j_0 \quad (5)$$

El cálculo de los coeficientes de aproximación y de detalle se aborda mediante el Algoritmo Piramidal de Mallat, también conocido como Análisis Multirresolución [2]. Para analizar la señal en diferentes escalas se usan filtros digitales con frecuencias de corte distintas. Estos filtros dividen a la señal en cada nivel de descomposición en dos categorías: componentes de alta frecuencia y componentes de baja frecuencia. En la mayoría de señales son las componentes de baja frecuencia las que ofrecen la mayor parte de información de una señal, por lo que estas componentes serán las que se correspondan con los coeficientes de aproximación. Por el contrario, las componentes de alta frecuencia se corresponden con los coeficientes de detalle, otorgándole a la señal las características más particulares. Cabe destacar que, en cada nivel de descomposición será necesario realizar un submuestreo para que no se aumente al doble el número de muestras original. Un esquema de representación de este algoritmo puede verse en la fig. 1.

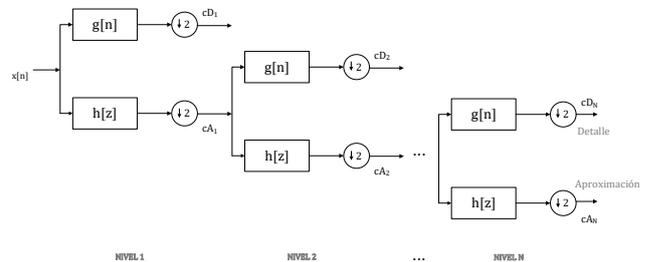


Fig. 1. Esquema de descomposición multinivel Wavelet.

**III. METODOLOGÍA. USO DE LA TRANSFORMADA WAVELET EN ALGORITMOS DE REDUCCIÓN DE RUIDO**

**A. Formulación del problema**

Antes de comenzar con el estudio de los diferentes algoritmos, debemos conocer cómo actúan éstos sobre la señal y el procedimiento general que siguen.

Un planteamiento común del problema es el siguiente. Supongamos que tenemos  $n$  muestras de una señal ruidosa de la forma:

$$s(n) = f(n) + \sigma n(n) \quad (6)$$

donde  $f(n)$  es la señal de interés,  $n(n)$  es ruido blanco Gaussiano independiente e idénticamente distribuido (i.i.d.)  $N(0, \sigma^2)$  añadido a la señal e independiente de la misma y  $\sigma$  hace referencia al nivel de dicho ruido.

Las técnicas de reducción de ruido estudiadas involucran los siguiente tres pasos:

1. Descomposición. Obtención de los coeficientes wavelet de la señal ruidosa haciendo uso del algoritmo Piramidal de Mallat. Deberá elegirse la Wavelet Madre y el número de niveles de descomposición.

$$y_{j,k} = DWT\{s(n)\} = w_{j,k} + \sigma v_{j,k} \quad (7)$$

donde

$$w_{j,k} = DWT\{f(n)\} \quad (8)$$

$$v_{j,k} = DWT\{n(n)\} \quad (9)$$

$j$  denota el nivel de descomposición y  $k$  es el índice de cada coeficiente en el nivel  $j$ .

Los coeficientes wavelet de la señal observada pueden ser considerados como una versión ruidosa de los coeficientes wavelet de la señal original. Además, la expresión (7) denota que puede reformularse el problema de recuperar la función  $f$  como aquel en el que se quieren recuperar los coeficientes de mayor amplitud, es decir, aquellos que son relativamente más fuertes que el fondo ruidoso. Este enfoque en el que cada coeficiente se compara con un umbral determinado para decidir si constituye parte deseable de la señal se denomina *wavelet thresholding*.

2. Elección del umbral y eliminación de coeficientes. La idea básica es eliminar los coeficientes que están por debajo de un cierto umbral  $\lambda$ . La elección de este umbral es la diferencia más significativa en la mayoría de los algoritmos propuestos. Para la eliminación de coeficientes se pueden aplicar dos versiones, *hard threshold* o *soft threshold*, las cuales vienen dadas respectivamente por las siguientes expresiones:

$$\delta_H(d_{j,k}, \lambda) = \begin{cases} 0, & \text{si } |d_{j,k}| \leq \lambda \\ d_{j,k}, & \text{si } |d_{j,k}| > \lambda \end{cases} \quad (10)$$

$$\delta_S(d_{j,k}, \lambda) = \begin{cases} 0, & \text{si } |d_{j,k}| \leq \lambda \\ \text{sgn}(d_{j,k})(|d_{j,k}| - \lambda), & \text{si } |d_{j,k}| > \lambda \end{cases} \quad (11)$$

3. Reconstrucción de la señal mediante la Transformada Wavelet Inversa.

$$\hat{f}(t) = IDWT\{Z\} \quad (12)$$

siendo

$$Z = \delta(y, \lambda) \quad (13)$$

Como ya hemos mencionado, los algoritmos que usan técnicas wavelets, como los que describiremos en la siguiente sección, tratan de estimar el valor óptimo de  $\lambda$ . Sin embargo, el primer paso en estos algoritmos es estimar el nivel de ruido,  $\sigma$ . Una

buena estimación del nivel de ruido fue propuesta por Donoho y Johnstone [3] y puede verse en la expresión (14).

$$\hat{\sigma} = \frac{\text{median}(\{|d_{j,k}\}| : k = 0, 1, \dots, 2^{j-1} - 1)}{0.6745} \quad (14)$$

## B. Análisis de algoritmos

Una vez expuesto el procedimiento común que siguen los métodos para reducción de ruido basados en wavelet, podemos estudiar la forma en la que cada algoritmo obtiene el umbral  $\lambda$ .

### 1. Minimax

Este algoritmo fue desarrollado por David Donoho y Iain Johnston en [4].

$$\lambda_M = \begin{cases} 0, & n \leq 32 \\ 0.3936 + 0.1829 \left( \frac{\log(n)}{\log 2} \right), & n > 32 \end{cases} \quad (15)$$

siendo  $n$  la longitud de la señal.

### 2. Universal Threshold

Este algoritmo, que surgió como alternativa al método anterior, es también conocido como *VisuShrink* y fue propuesto por los mismos autores [5].

$$\lambda_U = \sigma \sqrt{2 \log_2 n} \quad (16)$$

donde  $\sigma$  es la desviación estándar del ruido y  $n$  es la longitud de la señal.

### 3. Bayes Shrink

La mayor parte de los métodos propuestos a lo largo de la historia, tras las aportaciones de Donoho y Johnstone, son demasiado generales, es decir, no tienen en cuenta las propiedades específicas de cada señal y suavizan demasiado las señales a las que son aplicados. Por tanto, es necesario encontrar un umbral que no sea tan agresivo como los anteriores. En este contexto se sitúa el algoritmo Bayes Shrink.

$$\lambda_M = \begin{cases} \frac{\sigma^2}{\sigma_f}, & \text{si } \sigma^2 < \sigma_s^2 \\ \max(|y_{j,k}|), & \text{si } \sigma^2 > \sigma_s^2 \end{cases} \quad (17)$$

siendo  $\sigma$  la desviación estándar del ruido,  $\sigma_s^2$  la varianza de la señal contaminada y  $\sigma_f$  la desviación estándar de la señal pura. Estas dos últimas pueden calcularse mediante las expresiones (18) y (19), respectivamente.

$$\sigma_s^2 = \frac{1}{n} \sum_{j,k} y_{j,k}^2 \quad (18)$$

$$\sigma_f = \sqrt{\max(\sigma_s^2 - \sigma^2, 0)} \quad (19)$$

### 4. Normal Shrink

Este método, que fue desarrollado en [6], es una modificación del método anterior, y consiste en introducir un factor  $\beta_j$  en el umbral obtenido para cada nivel de resolución por el método *Bayes Shrink*.

$$\beta_j = \sqrt{\log \left( \frac{n}{j} \right)} \quad (20)$$

El umbral quedará, por tanto, de la siguiente forma:

$$\lambda_{NS} = \beta \frac{\sigma^2}{\sigma_f} \quad (21)$$

## 5. Modified Shrink

Este algoritmo es también una modificación del umbral propuesto por el método *Bayes Shrink*.

$$\beta_j = \sqrt{\log \frac{n}{2j}} \quad (22)$$

$$\lambda_{MS} = \beta \frac{\sigma^2}{\sigma_f} \quad (23)$$

## 6. Threshold dependent on Signal to Noise Ratio

Este algoritmo ha sido propuesto durante la realización de este Proyecto, con el objetivo de encontrar un umbral variable que se adapte a la señal y al nivel de ruido presente en cada nivel de resolución. El parámetro que mejor define la calidad de una señal es la relación señal a ruido o SNR. Por este motivo, hemos desarrollado un algoritmo en el que la elección del umbral para cada nivel depende de dicho parámetro. Esta técnica parte del umbral propuesto por el algoritmo *Universal Threshold* y lo modifica mediante la sustracción de una cantidad  $m$ . Nuestro objetivo es encontrar un umbral no tan agresivo y elevado como el obtenido mediante la técnica anterior.

$$\lambda_{SNR_j} = \sigma \sqrt{2 \log n} - m \quad (24)$$

$$m = \frac{pSNR_j}{10} \quad (25)$$

El valor de  $m$  ha sido calculado de forma empírica, a partir de resultados y comprobaciones experimentales.

## IV. DESARROLLO EXPERIMENTAL Y ANÁLISIS DE RESULTADOS

Para realizar un estudio y valoración del rendimiento de estas técnicas en señales de voz trabajaremos con señales de la Base de Datos Aurora 2 de dígitos conectados, la cual se preparó como una contribución al Grupo de Trabajo ETSI STQ-AURORA DSR en algoritmos de reconocimiento de voz en condiciones ruidosas y que se ha puesto a disposición del público a través de la Asociación Europea de Recursos Lingüísticos (ELRA). Ésta añade ruido a las señales en una gama de relación señal a ruido que se extiende de 0 a 20 dB y está compuesta por grabaciones de hombres y mujeres muestreadas a 8 kHz. El ruido añadido a las señales de las que disponemos es una selección de 4 ruidos diferentes del mundo real: tren de cercanías, murmullo de gente, coche y sala de exhibiciones. Éstos han sido elegidos por representar los escenarios de aplicación más probables.

En cuanto a señales de naturaleza sísmica se trabajará con sismogramas reales de tres componentes, extraídos del Etna, volcán situado en la costa este de la isla italiana de Sicilia, entre las provincias de Mesina y Catania.

### A. Señales de voz

Para evaluar el comportamiento de las técnicas propuestas en señales de voz se ha realizado un test MOS (*Mean Opinion Score*) de inteligibilidad. La calidad del audio ha sido evaluada directamente, lo que se conoce como *Absolute Category Rating* (ACR). Los audios han sido calificados con valores entre 1 y 5, siendo 5 'Excelente' y 1 'Malo'. El test ha sido completado por 16 oyentes de diferentes edades y sexo, lo que nos proporciona una mayor validez en los resultados. A continuación, se muestran las gráficas obtenidas para cada

ruido. En ellas se representa la inteligibilidad en función de la SNR, para cada una de las técnicas que han sido aplicadas.

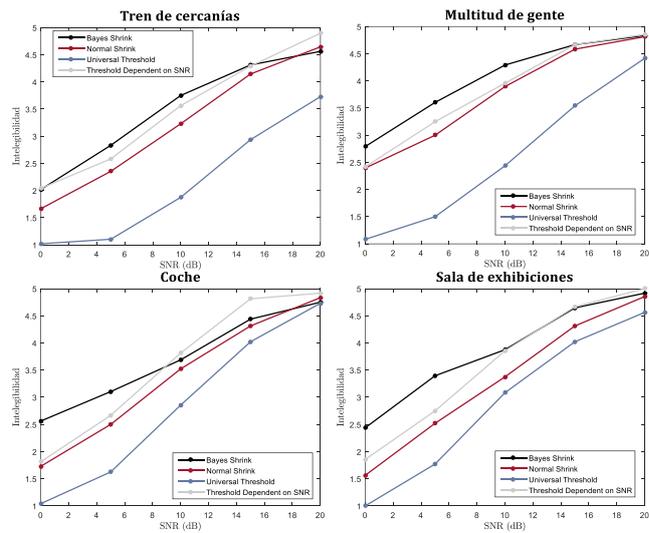


Fig. 2. Análisis del comportamiento de cada técnica en señales contaminadas con los cuatro tipos de ruido. Se representa la inteligibilidad de las señales frente al valor de SNR de éstas.

En la fig. 2 puede comprobarse que las técnicas que mejor funcionan son *Bayes Shrink*, *Threshold dependent on SNR* y *Normal Shrink*. A diferencia del algoritmo *Universal Threshold*, estas tres técnicas tienen un comportamiento relativamente bueno cuando la SNR es baja. Esto se debe, a que la técnica *Universal Threshold* es demasiado agresiva y da lugar a la eliminación de demasiados coeficientes, originando un sonido metálico desagradable al oído humano. Además, es importante destacar que las técnicas que mejor funcionan son aquellas que tienen en cuenta las características de la señal y que obtienen un umbral variable en cada nivel de descomposición, dando lugar a resultados más precisos.

Para obtener una visión gráfica del comportamiento de estas técnicas, a continuación, se representa una comparación de la señal ruidosa original, de la señal limpia y de la señal obtenida tras aplicar dos de las técnicas mencionadas. Esto nos permitirá comprobar la cantidad de ruido que ha sido eliminado y se podrá examinar el parecido existente entre la señal limpia obtenida y la señal sin ruido. Se mostrará la misma comparación para los espectrogramas de cada señal, con el fin de examinar su contenido en frecuencia.

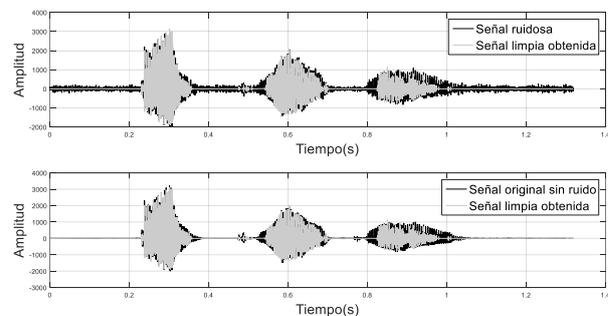


Fig. 3. Análisis de la señal 'eight three three' pronunciada por el locutor FAP tras aplicar el algoritmo *Universal Threshold* (soft).

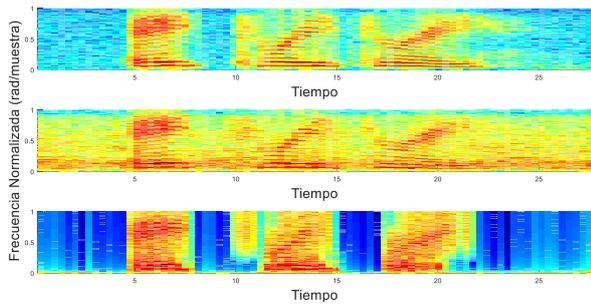


Fig. 4. Análisis del espectro de la señal ‘eight three three’ pronunciada por el locutor FAP tras aplicar el algoritmo *Universal Threshold (soft)*.

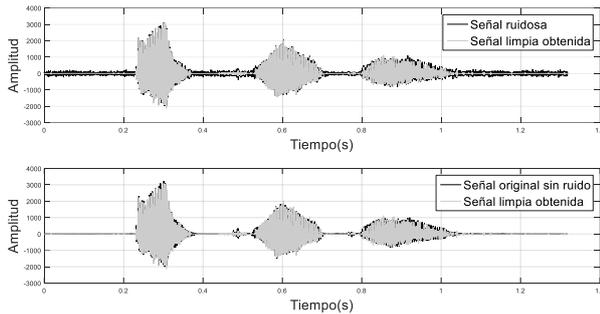


Fig. 5. Análisis de la señal ‘eight three three’ pronunciada por el locutor FAP tras aplicar el algoritmo *Threshold dependet on SNR (soft)*.

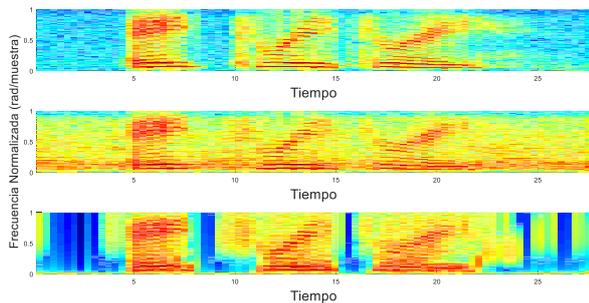


Fig. 6. Análisis del espectro de la señal ‘eight three three’ pronunciada por el locutor FAP tras aplicar el algoritmo *Threshold dependet on SNR (soft)*.

Puede comprobarse gráficamente el comportamiento tan agresivo de la técnica *Universal Threshold* en comparación con la técnica *Threshold dependet on SNR*. Tanto en la señal en el dominio del tiempo como en los espectrogramas puede apreciarse la severa reducción de coeficientes, llegando incluso a eliminar parte significativa de la señal.

Cabe destacar que, para todos los casos, la wavelet madre usada para la descomposición wavelet ha sido la ‘db8’ de la familia Daubechies, la cual implica que tengamos filtros de 16 coeficientes.

**B. Señales sísmicas**

Para evaluar el comportamiento de las técnicas propuestas en señales sísmicas, se ha usado un algoritmo de detección automática del *picking* desarrollado por miembros del Departamento de Teoría de la Señal, Telemática y Comunicaciones de la Universidad de Granada.

Una correcta detección de los tiempos de llegada de las ondas P y S, permite clasificar los distintos eventos sísmicos, así como realizar un estudio de los fenómenos que se producen en el interior de un volcán y de su estructura interna. En la mayoría de los casos, dicha detección se realiza de forma manual, siendo un experto en la materia el encargado de realizar el proceso. Sin embargo, este proceso es lento y puede ocasionar fallos debidos a la imprecisión humana. Por este motivo, en los últimos años, la comunidad científica se ha centrado en desarrollar algoritmos de *picking* automático de ondas sísmicas, que permitan trabajar con grandes bases de datos y adoptar el mismo criterio en la obtención de tiempos.

En nuestro caso, se comparará la diferencia entre el *picking* manual y el obtenido con el algoritmo de detección automática mencionado, para señales limpias tras la aplicación de las técnicas de reducción de ruido propuestas. Nuestro objetivo es que la diferencia de tiempos de *picking* entre los proporcionados por el algoritmo automático y los que proporciona el catálogo manual sea lo menor posible, pues de esta forma se estará asegurando la validez y el correcto funcionamiento de nuestras técnicas. Además, desde un punto de vista físico, el hecho de que los tiempos de *picking* se acerquen a los proporcionados por el catálogo manual, nos permite trabajar de forma más segura con este tipo de señales, pues aumentaremos la precisión en la adquisición de tiempos.

Así pues, el algoritmo ha sido usado de la siguiente forma. Para aquellas señales en las que la diferencia entre el *picking* manual y el automático sea mayor a 0.5 segundos, el algoritmo automático no obtiene un valor que se aproxime al valor manual debido al elevado nivel de ruido presente en la señal, por lo que será conveniente aplicar alguna técnica de reducción de ruido en dichas señales.

En el caso de la onda P, de las 2146 señales que forman parte de la base de datos, para 222 señales se cumple que la diferencia entre el *picking* manual y el automático es superior a 0.5 segundos. Para la onda S, un total de 742 cumplen lo anterior. Por tanto, aplicaremos las técnicas de reducción de ruidos a las señales mencionadas.

A continuación, se muestran dos diagramas de barras que representan el porcentaje de señales en las que mejora el *picking* de la onda P y de la onda S después de haberles aplicado las técnicas de reducción de ruido estudiadas.

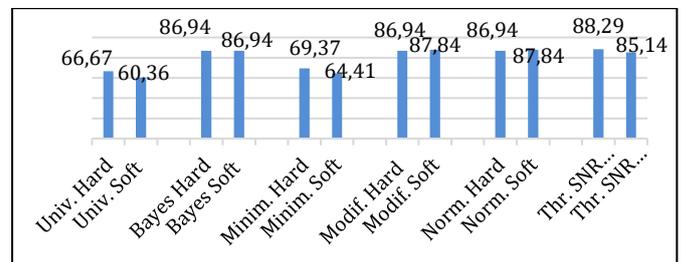


Fig. 7. Porcentaje de señales que mejoran el *picking* de la onda P tras la aplicación de las técnicas de reducción de ruido con wavelet.

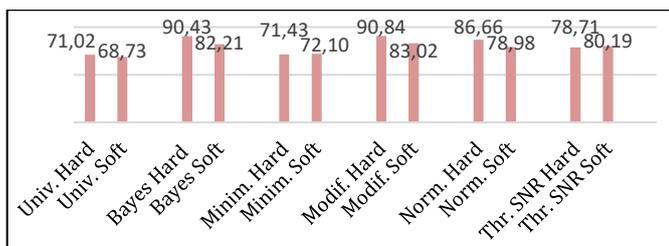


Fig. 8. Porcentaje de señales que mejoran el picking de la onda S tras la aplicación de las técnicas de reducción de ruido con wavelet.

Los resultados obtenidos muestran que los algoritmos *Minimax* y *Universal Threshold* son los que peores resultados ofrecen. Esto se debe a que no tienen en cuenta las características de la señal para obtener el valor del umbral, tal y como ocurría en señales de voz. Estas técnicas son demasiado agresivas y eliminan componentes importantes, lo que dificulta la correcta detección de los tiempos de llegada de la onda P y de la onda S.

A continuación, con el objetivo de ver el comportamiento de las técnicas en terremotos volcánico-tectónicos de forma gráfica, se muestra un análisis comparativo. Se mostrará la señal ruidosa en sus tres componentes y la señal limpia obtenida tras la aplicación de una de las técnicas. En este caso se muestra el resultado del algoritmo *Threshold dependent on SNR*, por ser el que mejores resultados ofrece en general.

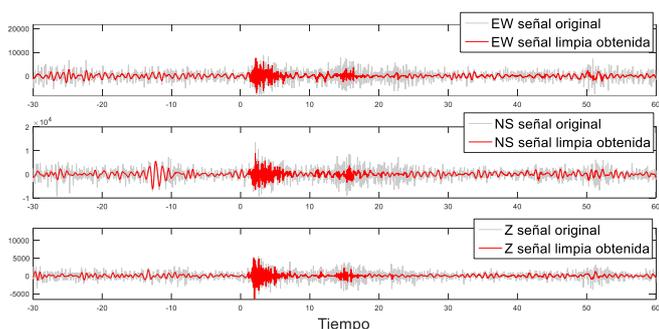


Fig. 9. Análisis del evento producido por el volcán Etna y registrado por la estación EPLC. En la parte superior se representa la componente E-W original y la obtenida tras la aplicación del algoritmo *Threshold dependet on SNR* (soft). En el centro aparece la componente N-S y en la parte inferior la componente Z.

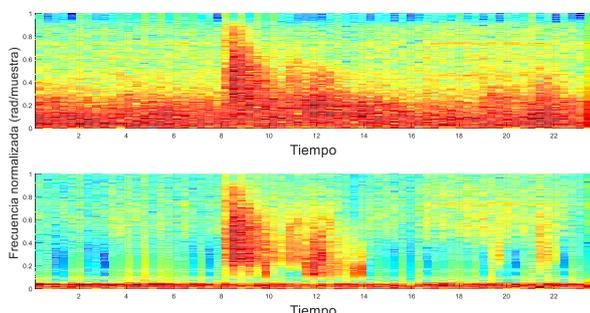


Fig. 10. Espectro de la componente Z del evento producido por el volcán Etna y registrado por la estación EPLC. En la imagen superior se muestra el espectro original y en la imagen inferior el obtenido tras la aplicación del algoritmo *Threshold dependet on SNR* (soft).

En este caso se ha selección la wavelet madre ‘*coif5*’ de la familia Coiflets para la descomposición. Esta implica que

tengamos filtros de 30 coeficientes.

## V. CONCLUSIONES

El análisis de resultados es el proceso clave para valorar el funcionamiento de las técnicas implementadas y para analizar si los objetivos establecidos se han cumplido satisfactoriamente.

En general, las técnicas tienen un efecto positivo en ambos tipos de señales, pues consiguen reducir el ruido. Es cierto que no siempre funcionan correctamente, pues estamos trabajando con señales reales, las cuales difieren de un comportamiento ideal. No obstante, el porcentaje de fallos es bastante reducido y depende de la técnica empleada. Además, el comportamiento de cada algoritmo es similar en los dos tipos de señales estudiados, lo que significa que nuestro objetivo de encontrar técnicas que se adaptaran a la naturaleza de cada señal ha sido cumplido.

En cuanto al trabajo futuro se proponen dos líneas que podrían significar una mejora en el trabajo realizado. La primera de ellas consiste en desarrollar un algoritmo que detecte de forma automática la wavelet madre que mejor se adapta a cada señal y que origina una mejor reconstrucción. La segunda línea consiste en realizar un estudio del contenido frecuencial del ruido presente en la señal y en base a éste elegir una descomposición en niveles a partir de la rama de filtros paso-baja o de filtros paso-alta.

## REFERENCIAS

- [1] Bonnefroy-Claudet, S., Cotton, F. y Bard, P. “The nature of noise wavefield and its applications for site effects studies”, University Joseph Fourier, France, 2006.
- [2] Mallat, S. G. y Santos, E., “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, IEEE Transactions on pattern analysys and machine intelligence”, vol.2, no. 7, Julio 1989.
- [3] Donoho, D. L. y Johnstone, I. M. “Adapting to Unknown Smoothness via Wavelet Shrinkage”, Department of Statistics, Stanford University, July 1994.
- [4] Donoho, D. L. y Johnstone, I. M. “Minimax estimation via wavelet shrinkage”, (1998) The Annals of Statistics, 26(3), 879-921.
- [5] Donoho, D. L. y Johnstone, I. M. “Ideal spatial adaptation by wavelet”, (1994) Biometrika, 81(3), 425-455.
- [6] Kaur, L., Gupta, S. y Chauhan, R. C., “Image Denoising using Wavelet Thresholding, Dept. of Computer Science and Engineering, Sant Longowal Institute of Engineering and Technology, Longowal, India.



**Laura Gámiz Pérez**, nacida en Granada el 27 de Diciembre de 1995, graduada en Ingeniería de Tecnologías de Telecomunicación (2013-2017) por la Universidad de Granada.

# Deep neural networks based multichannel speech enhancement

Author: Juan Manuel Martín Doñas, e-mail: mdjuamart@ugr.es

Tutor: Ángel Manuel Gómez García, e-mail: amgg@ugr.es

Tutor: Antonio Miguel Peinado Herreros, e-mail: amp@ugr.es

Master's degree: Telecommunication Engineering

Dept. of Signal Theory, Telematics and Communications  
University of Granada

**Abstract**—Speech communications in real-world scenarios need high performance enhancement algorithms to address the distortions that can degrade the intelligibility and quality of the speech signal. Current portable devices usually integrate multiple microphones that can conveniently be exploited to improve the signal quality. In this paper we present a dual-microphone speech enhancement approach suitable for smartphones with primary (front) and reference (back) microphones. Our proposal is based on the use of deep neural networks which are able to obtain a non-linear mapping function between noisy and clean speech signals. We explore two different architectures: a feedforward deep neural network (DNN) with temporal context and a gated recurrent unit (GRU) recurrent neural network (RNN). The proposed system is evaluated under different acoustic conditions in close- and far-talk device positions. A comparison with other single- and dual-channel approaches shows that our proposal obtains the best performance in terms of perceptual quality.

**Keywords**—Speech Enhancement, Dual-Channel, Deep Neural Network, Recurrent Neural Network, Smartphones

## I. INTRODUCTION

MOBILE phone communications, both human-human or human-machine, are a key aspect of our current society, especially due to the emergence of smartphones and other similar devices. In real-world environments, the speech signal captured by a smartphone might be corrupted by different types of distortions such as background noise, interference, echoes and reverberation, that degrade the signal quality and can reduce its intelligibility. Indeed, these distortions also deteriorate the performance of speech coding and recognition systems. Hence, it is desirable to develop high performance speech enhancement algorithms to improve the intelligibility and perceptual quality of the degraded signals [1].

Most of the classical single-channel noise reduction algorithms based on the estimation of the speech spectral amplitude can be classified into spectral subtraction, Wiener filtering-based, and statistical model-based or Bayesian methods [2]. The latter type of methods was first proposed in the Ephraim and Malah works [3], [4], which were later improved by other suggested estimators, as the well-known optimally-modified log-spectral amplitude (OM-LSA) estimator [5] proposed by Cohen *et al.* These techniques usually rely on accurate estimations of the noise power spectral density (PSD). The improved minima controlled recursive averaging (IMCRA) noise estimator [6], applied in the OM-LSA algorithm, is an example of this type of techniques. Despite their good performance, these single-channel methods are quite

limited when performing in real-life scenarios because of the existing trade-off between noise reduction and introduced speech distortion.

Currently, mobile devices often integrate two or more microphones. This multi-channel information can be exploited to improve the performance of the speech enhancement algorithms. Beamforming is a kind of multi-microphone approaches widely used for speech enhancement purposes that can reduce noise by applying a spatial filtering. In addition, it is a common practice to apply a (single-channel) post-filter at the output of the beamformer to improve its performance, e.g. [7]. Nevertheless, in devices with only two microphones, particularly when they are located on the device's front and back sides (for example, to try to capture separately the speech signal and the environmental noise), the performance of beamforming can be severely degraded [8]. Thus, in [9], Jeub *et al.* proposed a noise PSD estimator and a noise reduction algorithm specially tailored for dual-microphone mobile phones when they are used in close-talk position. This method explicitly exploits the power level difference (PLD) of the speech signal between the microphones. Nelke *et al.* developed an alternative dual-channel method [10] to estimate the noise PSD in a mobile phone, used in far-talk conditions, based on speech presence probability (SPP) and coherence properties of the target signal and the background noise. Although these methods exploit the dual-channel information, they are still based on assumptions about the speech and noise signals that may not be realistic in practice, leading to a limited performance.

In recent years, deep neural networks (DNNs) have become very popular because of their ability to deal with highly complicated tasks where classical approaches are limited. This is due to their capability to model high level abstractions of data and tackle the difficult non-linear relationships as they exist between noisy and clean speech signals in real-world environments. Deep learning methods are widely used in speech processing, mainly in speech recognition but also to carry out other tasks such as synthesis, coding and speech enhancement. In particular, DNNs for single-channel speech enhancement have been used in different approaches: modeling the mapping function between noisy and clean signals [11], feature enhancement for automatic speech recognition (ASR) [12], estimation of ideal binary mask (IBM) [13] or ideal ratio mask (IRM) [14], [15], etc. In connection with our proposal, Xu *et al.* [16] proposed a DNN regression-

based single-channel speech enhancement system trained on different noise types and using noise-aware training (NAT) to increase the generalization capability of the network to unseen environments during the training phase, showing a noticeable improvement in comparison with previous DNN-based enhancement approaches.

On the other hand, a number of ASR systems exploit multi-channel signals at the front-end subsystem to increase the recognition rate. Also, challenges like CHiME-3 [17] and CHiME-4 [18] have been proposed to increase use of multi-channel ASR in portable devices. In addition, different deep learning approaches have been followed to strengthen ASR systems against noise in multi-channel environments: using a conventional beamformer along DNN-based single-channel post-filtering [19], estimation of the beamformer weights [20], or time-frequency masks [21], [22] employed to calculate the beamforming weights. Other works [23], [24], [25], [26] use multi-channel features as input of their acoustic modeling DNN system for recognition purposes. Indeed, we recently proposed in [27] a dual-channel DNN-based noise estimator for feature compensation in speech recognition systems. In [28] and [29], filter-and-sum beamformers are developed by using supervised learning of the filter coefficients with convolutional networks integrated in ASR systems. However, in all these works the final objective is speech recognition and there is not reconstruction of the enhanced signal.

There are a few DNN-based works where multi-channel information is employed for estimation of time-frequency masks [30], [31], [32], [33] or multi-channel Wiener filters [34], [35] where the final objective is multi-channel speech enhancement or source separation. More specifically, in [36], a multi-channel DNN-based regression system is employed for binaural speech segregation which exploits the interaural level difference (ILD) between left and right channels.

In summary, in the context of speech enhancement for dual-microphone smartphones, while classical techniques like beamforming do not seem capable of achieving high performance, deep learning approaches have not been studied in depth when applied to this particular problem. To the best of our knowledge, a direct mapping of the noisy spectra, captured by a dual-microphone smartphone, to clean speech spectra using a DNN regression-based system has not been evaluated so far. This will be the main objective of this paper. We expect that the powerful abilities of DNNs allow us to better exploit the spatial information and improve the performance in comparison with the single-channel strategy. A feedforward deep neural network (DNN) with temporal context is first used. Then, we also study the use of a recurrent neural network (RNN) as a regressor to take advantage of the temporal dynamics present in the speech and noise signals.

The remainder of this paper is organized as follows. In Section 2, the dual-channel DNN-based enhancement approach and its different modules are described. Section 3 is divided in three subsections. First, the database used in the experiments is outlined. Then, the experimental framework and parameters of the system are described and, finally, the enhancement results are obtained and compared with those from other state-of-the-art techniques. Final conclusions are summarized in Section 4.

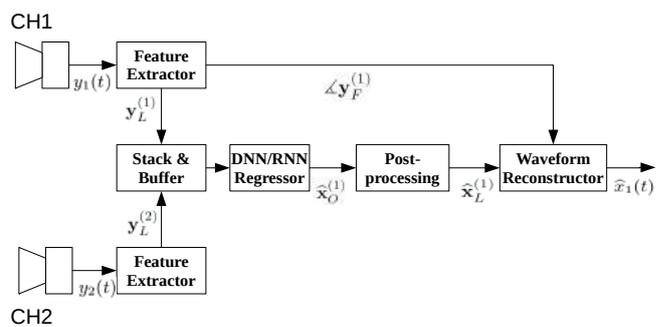


Fig. 1. Block diagram of the proposed dual-channel DNN/RNN-based enhancement system.

## II. DUAL-CHANNEL DNN-BASED ENHANCEMENT APPROACH

A block diagram of our system is depicted in Fig. 1. A DNN is trained for mapping the dual-channel noisy speech features to clean features at the first channel. The system is constructed using the following modules in cascade: feature extractor, DNN regressor, post-processor and waveform reconstructor.

First, the feature extractor transforms the raw speech signals  $y_1(t)$  and  $y_2(t)$  to obtain the input vectors used by the DNN regressor. Log-power spectral features are chosen. A Short-Time Fourier Transform (STFT) with overlapped windows is applied over the speech signal. The log-power spectrum (LPS)  $y_L^{(i)}$  of each temporal frame is then calculated as

$$\mathbf{y}_L^{(i)} = \log \left| \mathbf{y}_F^{(i)} \right|^2, \quad (1)$$

where  $\mathbf{y}_F^{(i)}$  is a vector with the first  $N+1$  complex coefficients of the  $2N$  DFT values of the speech frame, and  $i$  refers to the channel (i.e.  $i = 1, 2$ ). The natural logarithm is applied element-wise. Feature vectors of dimension  $2(N+1)$  are shaped by stacking the LPS features from both channels. Mean and variance normalization are jointly applied to the components of the LPS feature vectors from both channels to keep the power level differences unaltered. The mean ( $\mathbf{m}$ ) and standard deviation ( $\mathbf{d}$ ) vectors are obtained from the noisy LPS vectors of the training dataset. Also, we apply the same normalization to the target vectors during the DNN training phase.

The mapping function between dual-channel noisy and clean speech features is implemented by a deep neural network previously trained using normalized noisy and clean features. Two different architectures are compared. First, a feedforward deep neural network (DNN), where the input of the DNN is a context window of  $\tau$  look-backward and look-forward frames centered on the current dual-channel frame. Therefore, the input dimension is  $2(2\tau+1)(N+1)$  and the output dimension is  $N+1$ . To improve the system by taking advantage of the temporal dynamics of speech, a recurrent neural network (RNN) is also used as a second architecture. Gated recurrent units (GRU) [37] are chosen due to their simplicity and similar performance to the well-known long short-term memory (LSTM) units. In this case, the input vectors have dimension  $2(N+1)$  each. The GRU approach uses information of all previous frames and only a context of future frames, known as *look ahead*. Block diagrams of the

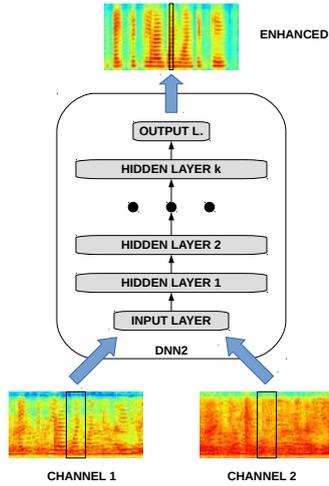


Fig. 2. Block diagram of the DNN-based regressor.

DNN-based regressor and RNN-based regressor are depicted in Fig. 2 and Fig. 3, respectively.

A problem with the feedforward DNN regressor is that the output is over-smoothed compared to the target clean features. A straightforward solution is to apply a post-processing stage to the enhanced features. Frequency-independent Global Variance Equalization (GVE) is adopted as post-processing technique [16]. GVE estimates a multiplicative factor for the LPS enhanced features from the relationship between the global variance of the enhanced and reference features of the training dataset. After applying GVE and denormalization, the enhanced LPS,  $\hat{\mathbf{x}}_L^{(1)}$ , is obtained as

$$\hat{\mathbf{x}}_L^{(1)} = \hat{\mathbf{x}}_O^{(1)} \cdot \eta \cdot \mathbf{d} + \mathbf{m}, \quad (2)$$

where  $\hat{\mathbf{x}}_O^{(1)}$  is the (normalized) output enhanced feature vector from the DNN regressor, and  $\eta$  is the GVE factor obtained as

$$\eta = \sqrt{\frac{GV_{ref}}{GV_{est}}}, \quad (3)$$

where  $GV_{ref}$  and  $GV_{est}$  are the frequency-independent global variance of the reference and estimated features, respectively.

Finally, the enhanced spectrum is obtained by using the enhanced LPS and the noisy phase  $\angle \mathbf{y}_F^{(1)}$  at the first channel as

$$\hat{\mathbf{x}}_F^{(1)} = e^{\frac{\hat{\mathbf{x}}_L^{(1)}}{2}} e^{j\angle \mathbf{y}_F^{(1)}}. \quad (4)$$

Then, the overlap-add method is applied to synthesize the waveform of the speech signal  $\hat{x}_1(t)$ .

### III. EXPERIMENTAL SETUP AND RESULTS

#### A. Databases

The AURORA2-2C-CT/FT (AURORA2 - 2 Channels - Close-Talk/Far-Talk) [38] are two synthetic noisy speech databases that contain speech signals recorded with a two-microphone smartphone in different environments and in two different configurations: close-talk (when the loudspeaker of the smartphone is placed at the ear of the user) and far-talk (when the user holds the device at a certain distance from

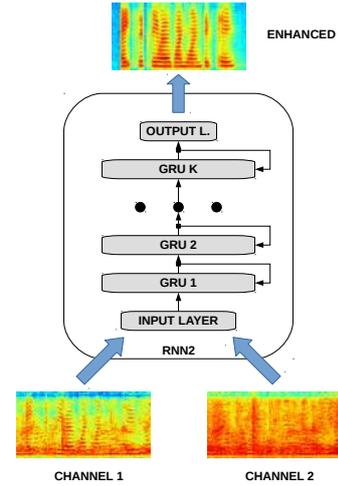


Fig. 3. Block diagram of the RNN-based regressor with GRU units.

her/his face). The clean speech signals of the AURORA2 database [39] are transformed to dual-channel signals according to the relative acoustic path between both microphones in the device, and dual-channel environmental noise is added at different SNR levels to simulate various scenarios.

Both databases (close-talk and far-talk) are arranged in three datasets: training, validation and test dataset. Four environments or noise types are used in the training and validation sets (bus, babble, car and pedestrian street). The test set is divided into two subsets: set A and set B. Set A employs the same four noise types as in the training, while in set B four different noises are chosen (cafe, street, bus station and train station) in order to study the performance of the approaches to unseen conditions during the training stage. Six different SNR levels from -5 dB to 20 dB are evaluated. For each database, a total of  $6 \text{ SNRs} \times 4 \text{ noises} \times 345 = 8280$  utterances form the training set. The validation set consists of  $6 \text{ SNRs} \times 4 \text{ noises} \times 40 = 960$  utterances while each subset in the test set has  $6 \text{ SNRs} \times 4 \text{ noises} \times 1001 = 24024$  utterances.

#### B. Experimental Framework

The speech signals are sampled at 8 kHz. Frame length is set to 256 samples with 128 samples overlapping. A Hanning window is used for the STFT computation. The dimension of the LPS vector is set to 129 (i.e.  $N = 128$ ) and the stacked vector with both channels is a 258-dimensional vector.

The dual-channel feedforward DNN uses a temporal context of  $\tau = 2$ , so the final input vector dimension is 1290. Three hidden layers with 2048 rectifier linear units (ReLU) are employed while the output layer dimension is 129. Dropout regularization [40] is used in the hidden layers with a deactivation probability factor of 0.1. On the other hand, the GRU-RNN architecture consists of an input layer with 258 components, 3 recurrent layers of 800 units and a linear output layer of dimension 129. A *look ahead* of 4 subsequent frames is applied.

In order to optimize the weights and biases of the network, the ADAM method [41] is chosen as optimization algorithm

Table I  
PESQ RESULTS FOR TEST A SUBSET IN CLOSE-TALK CONDITION.

SNR (dB)	Noisy	LSA	PLD	DNN1	DNN2	RNN2
<b>20</b>	3.1293	3.3917	3.2354	3.5751	<b>3.6471</b>	3.5367
<b>15</b>	2.8599	3.1145	3.0034	3.4300	<b>3.5059</b>	3.4274
<b>10</b>	2.5673	2.8010	2.7671	3.2271	<b>3.3084</b>	3.2814
<b>5</b>	2.2496	2.4487	2.5204	2.9592	3.0538	<b>3.0900</b>
<b>0</b>	1.9271	2.0566	2.2385	2.5848	2.6976	<b>2.8233</b>
<b>-5</b>	1.6670	1.6751	1.9169	2.0671	2.1928	<b>2.4358</b>
<b>Avg.</b>	2.4000	2.5813	2.6136	2.9739	3.0676	<b>3.0991</b>

Table II  
PESQ RESULTS FOR TEST B SUBSET IN CLOSE-TALK CONDITION.

SNR (dB)	Noisy	LSA	PLD	DNN1	DNN2	RNN2
<b>20</b>	2.9545	3.2158	3.0977	3.3748	<b>3.4632</b>	3.3971
<b>15</b>	2.6672	2.9260	2.8646	3.1567	<b>3.2468</b>	3.2335
<b>10</b>	2.3504	2.6149	2.6187	2.8725	2.9762	<b>3.0229</b>
<b>5</b>	2.0136	2.2558	2.3514	2.5255	2.6676	<b>2.7687</b>
<b>0</b>	1.7145	1.8459	2.0387	2.0598	2.2625	<b>2.4122</b>
<b>-5</b>	1.5576	1.4167	1.7109	1.5375	1.7757	<b>1.9583</b>
<b>Avg.</b>	2.2096	2.3792	2.4470	2.5878	2.7320	<b>2.7988</b>

Table III  
PESQ RESULTS FOR TEST A SUBSET IN FAR-TALK CONDITION.

SNR (dB)	Noisy	LSA	SPPC	DNN1	DNN2	RNN2
<b>20</b>	3.1886	3.4078	2.7534	3.5788	<b>3.6203</b>	3.5410
<b>15</b>	2.9190	3.1295	2.6402	3.4316	<b>3.4781</b>	3.4292
<b>10</b>	2.6265	2.8106	2.4861	3.2282	3.2724	<b>3.2735</b>
<b>5</b>	2.3024	2.4583	2.2943	2.9650	3.0068	<b>3.0693</b>
<b>0</b>	1.9847	2.0826	2.0691	2.6153	2.6590	<b>2.8039</b>
<b>-5</b>	1.6925	1.7018	1.8334	2.1181	2.1804	<b>2.4058</b>
<b>Avg.</b>	2.4523	2.5984	2.3461	2.9895	3.0362	<b>3.0871</b>

Table IV  
PESQ RESULTS FOR TEST B SUBSET IN FAR-TALK CONDITION.

SNR (dB)	Noisy	LSA	SPPC	DNN1	DNN2	RNN2
<b>20</b>	2.9783	3.2464	2.7123	3.3486	3.4099	<b>3.4180</b>
<b>15</b>	2.6980	2.9546	2.5855	3.1194	3.1693	<b>3.2519</b>
<b>10</b>	2.3844	2.6342	2.4117	2.8417	2.8694	<b>3.0342</b>
<b>5</b>	2.0512	2.2707	2.2036	2.4843	2.5172	<b>2.7587</b>
<b>0</b>	1.7360	1.8533	1.9586	2.0276	2.0670	<b>2.3752</b>
<b>-5</b>	1.5465	1.4379	1.7355	1.5606	1.5788	<b>1.8470</b>
<b>Avg.</b>	2.2324	2.3995	2.2679	2.5637	2.6019	<b>2.7808</b>

due to its fast convergence. The learning rates applied during training are  $10^{-4}$  and  $3 \cdot 10^{-4}$  for the feedforward DNN and the RNN, respectively. The *early-stopping* regularization technique [42] is applied by means of the validation set to prevent overfitting. For the design, training and testing of the DNN systems we use *Tensorflow* [43], an open-source software library for machine learning developed by Google.

### C. Results

In this subsection, we evaluate the proposed approaches using an objective measure, the Perceptual Evaluation of Speech Quality (PESQ) [44], that is highly correlated with subjective evaluation scores and is widely used for measuring the speech signal quality. To calculate the PESQ score, the enhanced speech is compared with a clean reference. PESQ values are in the range from -0.5 (poor quality) to 4.5 (best quality).

The dual-channel proposals DNN2 and RNN2 are compared with the noisy speech signal at the first microphone (Noisy) as well as with the following techniques: the OM-LSA algorithm for single-channel enhancement (LSA) [6], the Power Level Difference (PLD) method of [9] for dual-channel smartphones in close-talk conditions, the SPP- and coherence-based (SPPC) method of [10] for far-talk conditions (using Wiener filtering), and the single-channel DNN-based enhancement system (DNN1) in [16]. For the latter system, we use the same feedforward neural network architecture and optimization method as in DNN2, while only the primary channel is fed to the DNN. A temporal context of  $\tau = 4$  plus a noise vector for noise-aware training (NAT) are used as inputs, resulting in the same input dimension as that from our DNN2.

Tables I and II show the PESQ results for test sets A and B, respectively, in close-talk conditions, while Tables III and IV show the results for the same sets but in far-talk conditions. Average PESQ results are obtained for each of the 6 different SNR levels from -5 dB to 20 dB. Also, the average PESQ scores for every technique considered in every set (A or B) and condition type are shown in Fig. 4 and 5

(close-talk and far-talk scenario, respectively). In these figures the 95% confidence intervals obtained over the measures are also indicated.

A comparison among the different approaches shows that the proposed dual-channel methods outperform the other considered techniques. In particular, the recurrent neural network system achieves the best average results over the tested conditions. The obtained results remark the ability of DNNs to deal with difficult tasks without any kind of assumptions on the signals. By comparing OM-LSA with the PLD algorithm (close-talk conditions) we can observe that using dual-channel information is especially useful when the power level difference between the speech signals at both channels is significant. In far-talk conditions, the SPPC method does not exhibit a performance improvement over the single-channel approach OM-LSA possibly due to the difficulty of estimating the noise, which is a crucial issue for Wiener filters. However, at low SNRs the SPPC system yields better results, especially at -5 dB for test set B, where only RNN2 performs better.

Regarding the DNN-based approaches, in close-talk conditions, the dual-channel feedforward DNN (DNN2) obtains better quality results than the single-channel one (DNN1). This is particularly true at low SNRs, where higher relative increments on PESQ scores are observed. This is a noteworthy result which confirms the usefulness of the second microphone for enhancing degraded signals. In addition, the evaluation of both DNN1 and DNN2 in test set B, with noises not seen during the training phase, shows that DNN2 yields higher gains on PESQ scores than DNN1, possibly because the secondary channel is a better noise reference than the NAT vector in close-talk conditions. When comparing both dual-channel systems, it can be observed that DNN2 only performs slightly better than RNN2 at higher SNRs, while at lower SNRs RNN2 shows higher PESQ values, especially under the unseen noises of test set B. This fact illustrates the importance of taking into account the temporal information during the enhancement procedure. The slight performance reduction of RNN2 at high SNRs may be explained by the long temporal context being implicitly employed, which could introduce

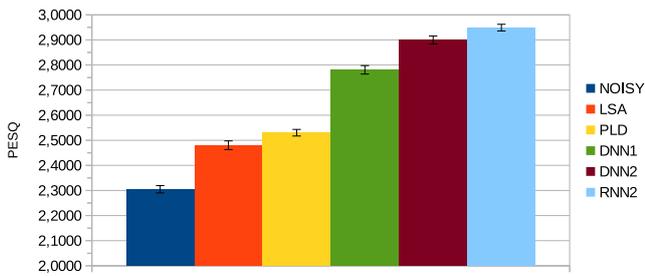


Fig. 4. Average PESQ results for close-talk scenario.

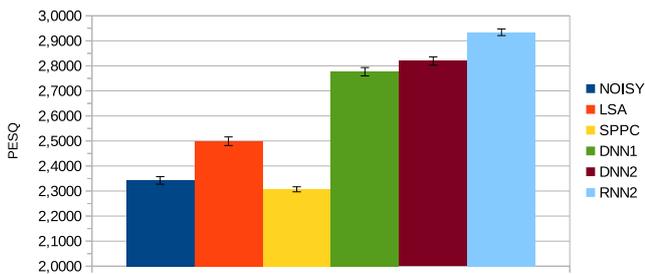


Fig. 5. Average PESQ results for the far-talk scenario.

some spectral over-smoothing. Despite that, informal listening shows us that, at high SNRs, a slight reduction of PESQ values is less harmful for the quality of the speech signal.

In far-talk conditions, the improvement achieved by DNN2 with respect to DNN1 is smaller than that of the close-talk case. This may be due to the fact that the power level difference between the speech signals at both microphones is reduced in the far-talk scenario and the noise reference at the secondary channel is not as good as in the close-talk case. Indeed, the far-talk condition is a more complex scenario where the clean speech acoustic path response between channels is more variable. Therefore, under unseen noise conditions during the training phase, the mapping function between noisy and clean speech is more difficult to model. However, when the dual-channel information is combined with the ability of the recurrent neural networks to deal with the signals dynamics, the performance is improved. In test set A, RNN2 provides, at mid SNRs, similar results than DNN2, while, at low SNRs, RNN2 outperforms DNN2. On the other hand, under the unseen conditions of test set B, RNN2 yields better PESQ scores in all cases, with a significant gain in the average PESQ result. This remarks again the importance of using not only the spatial but also the temporal information for the enhancement task.

#### IV. CONCLUSION

In this paper we have proposed a novel dual-channel DNN-based speech enhancement system for smartphones. The proposed approach exploits the dual-microphone information to develop a non-linear mapping function between noisy and clean speech spectra. Two different network architectures have been evaluated: DNN and RNN. Our proposals have been tested using the AURORA2-2C-CT/FT databases that contain noisy speech signals obtained with a dual-microphone smartphone in different noisy environments and configurations

(close-talk and far-talk positions). The results have shown that our proposals outperform other related approaches and that their simplicity make them especially suitable for smartphones.

#### ACKNOWLEDGMENT

This work has been supported by the Spanish MINECO/FEDER Project TEC2016-80141-P. This work has been published in the IEEE 19th International Workshop on Multimedia Signal Processing (MMSP) with the title *Dual-Channel DNN-based Speech Enhancement for Smartphones*.

#### REFERENCIAS

- [1] P. C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed. CRC Press, 2013.
- [2] M. Parchami, W. Zhu, B. Champagne, and E. Plourde, "Recent developments in speech enhancement in the short-time Fourier transform domain," *IEEE Circuits and Systems Magazine*, vol. 16, no. 3, pp. 45–77, 2016.
- [3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.
- [4] —, "Speech enhancement using a minimum mean-square error-log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 443–445, 1985.
- [5] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 81, no. 11, pp. 2403–2418, 2001.
- [6] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 466–475, 2003.
- [7] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proc. ICASSP*. IEEE, 1988, pp. 2578–2581.
- [8] I. Tashev, S. Mihov, T. Gleghorn, and A. Acero, "Sound capture system and spatial filter for small devices," in *Proc. InterSpeech*, 2008, pp. 435–438.
- [9] M. Jeub, C. Herglotz, C. Nelke, C. Beaugeant, and P. Vary, "Noise reduction for dual-microphone mobile phones exploiting power level differences," in *Proc. ICASSP*, 2012, pp. 1693–1696.
- [10] C. M. Nelke, C. Beaugeant, and P. Vary, "Dual microphone noise PSD estimation for mobile phones in hands-free position exploiting the coherence and speech presence probability," in *Proc. ICASSP*, 2013, pp. 7279–7283.
- [11] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder," in *Proc. InterSpeech*, 2013, pp. 436–440.
- [12] F. Weninger, J. Geiger, M. Wöllmer, B. Schuller, and G. Rigoll, "Feature enhancement by deep LSTM networks for ASR in reverberant multisource environments," *Computer Speech and Language*, vol. 28, no. 4, pp. 888–902, 2014.
- [13] Y. Wang and D. Wang, "Towards scaling up classification-based speech separation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 7, pp. 1381–1390, 2013.
- [14] A. Narayanan and D. Wang, "Ideal ratio mask estimation using deep neural networks for robust speech recognition," in *Proc. ICASSP*, 2013, pp. 7092–7096.
- [15] F. Weninger, H. Erdogan, S. Watanabe, E. Vincent, J. Le Roux, J. R. Hershey, and B. Schuller, "Speech enhancement with LSTM recurrent neural networks and its application to noise-robust ASR," in *Proc. LVA/IICA*, 2015.
- [16] Y. Xu, J. Du, L. R. Dai, and C. H. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE/ACM Transactions on Speech and Language Processing*, vol. 23, no. 1, pp. 7–19, 2015.
- [17] J. Barker, R. Marxer, E. Vincent, and S. Watanabe, "The third 'CHiME' speech separation and recognition challenge: Dataset, task and baselines," in *Proc. ASRU*, 2015, pp. 504–511.
- [18] E. Vincent, S. Watanabe, A. A. Nugraha, J. Barker, and R. Marxer, "An analysis of environment, microphone and data simulation mismatches in robust speech recognition," *Computer Speech and Language*, 2016.

- [19] T. Hori, Z. Chen, H. Erdogan, J. R. Hershey, J. Le Roux, V. Mitra, and S. Watanabe, "The MERL/SRI system for the 3rd CHiME challenge using beamforming, robust feature extraction, and advanced speech recognition," in *Proc. ASRU*, 2015, pp. 475–481.
- [20] X. Xiao, S. Watanabe, H. Erdogan, L. Lu, J. Hershey, M. L. Seltzer, G. Chen, Y. Zhang, M. Mandel, and D. Yu, "Deep beamforming networks for multi-channel speech recognition," in *Proc. ICASSP*, 2016, pp. 5745–5749.
- [21] J. Heymann, L. Drude, A. Chinaev, and R. Haeb-Umbach, "BLSTM supported GEV beamformer front-end for the 3rd CHiME challenge," in *Proc. ASRU*, 2015, pp. 444–451.
- [22] J. Heymann, L. Drude, and R. Haeb-Umbach, "Neural network based spectral mask estimation for acoustic beamforming," in *Proc. ICASSP*, 2016, pp. 196–200.
- [23] S. Renals and P. Swietojanski, "Neural networks for distant speech recognition," in *Proc. HSCMA*, 2014, pp. 172–176.
- [24] Y. Liu, P. Zhang, and T. Hain, "Using neural network front-ends on far field multiple microphones based speech recognition," in *Proc. ICASSP*, 2014, pp. 5542–5546.
- [25] I. Himawan, P. Motlicek, D. Imseng, and S. Sridharan, "Feature mapping using far-field microphones for distant speech recognition," *Speech Communication*, vol. 83, pp. 1–9, 2016.
- [26] K. H. Lee, T. G. Kang, W. H. Kang, and N. S. Kim, "DNN-based feature enhancement using joint training framework for robust multichannel speech recognition," in *Proc. InterSpeech*, 2016, pp. 3027–3031.
- [27] I. López-Espejo, A. M. Peinado, A. M. Gomez, and J. M. Martín-Doñas, "Deep neural network-based noise estimation for robust ASR in dual-microphone smartphones," in *Advances in Speech and Language Technologies for Iberian Languages: Third International Conference, IberSPEECH 2016*, vol. 10077 LNAI. Springer Verlag, 2016, pp. 117–127.
- [28] Y. Hoshen, R. J. Weiss, and K. W. Wilson, "Speech acoustic modeling from raw multichannel waveforms," in *Proc. ICASSP*, 2015, pp. 4624–4628.
- [29] T. N. Sainath, R. J. Weiss, K. W. Wilson, A. Narayanan, and M. Bacchiani, "Factored spatial and spectral multichannel raw waveform CLDNNs," in *Proc. ICASSP*, 2016, pp. 5075–5079.
- [30] P. Pertila and J. Nikunen, "Microphone array post-filtering using supervised machine learning for speech enhancement," in *Proc. InterSpeech*, 2014, pp. 2675–2679.
- [31] Y. Jiang, D. L. Wang, R. S. Liu, and Z. M. Feng, "Binaural classification for reverberant speech segregation using deep neural networks," *IEEE/ACM Transactions on Speech and Language Processing*, vol. 22, no. 12, pp. 2112–2121, 2014.
- [32] S. Araki, T. Hayashi, M. Delcroix, M. Fujimoto, K. Takeda, and T. Nakatani, "Exploring multi-channel features for denoising-autoencoder-based speech enhancement," in *Proc. ICASSP*. Institute of Electrical and Electronics Engineers Inc., 2015, pp. 116–120.
- [33] K. Niwa, Y. Koizumi, T. Kawase, K. Kobayashi, and Y. Hioka, "Pinpoint extraction of distant sound source based on DNN mapping from multiple beamforming outputs to prior SNR," in *Proc. ICASSP*, 2016, pp. 435–439.
- [34] S. Sivasankaran, A. A. Nugraha, E. Vincent, J. A. Morales-Cordovilla, S. Dalmia, I. Illina, and A. Liutkus, "Robust ASR using neural network based speech enhancement and feature simulation," in *Proc. ASRU*, 2015, pp. 482–489.
- [35] A. A. Nugraha, A. Liutkus, and E. Vincent, "Multichannel audio source separation with deep neural networks," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 24, no. 9, pp. 1652–1664, 2016.
- [36] N. Fan, J. Du, and L. R. Dai, "A regression approach to binaural speech segregation via deep neural network," in *Proc. ICSLP*, 2016, pp. 1–5.
- [37] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN Encoder-Decoder for statistical machine translation," 2014. [Online]. Available: <http://arxiv.org/abs/1406.1078>
- [38] I. López-Espejo, A. M. Gomez, J. A. González, and A. M. Peinado, "Feature enhancement for robust speech recognition on smartphones with dual-microphone," in *Proc. EUSIPCO*, 2014.
- [39] D. Pearce and H. Hirsch, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *Proc. ICSLP*, 2000.
- [40] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012. [Online]. Available: <http://arxiv.org/abs/1207.0580>
- [41] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference for Learning Representations*, San Diego, 2015, pp. 1–13.
- [42] L. Prechelt, "Early Stopping - But When?" in *Neural Networks: Tricks of the Trade*. Springer Berlin Heidelberg, 2012, pp. 53–67.
- [43] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: <http://download.tensorflow.org/paper/whitepaper2015.pdf>
- [44] "ITU-T. Rec. P.862, Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assesment of narrow-band telephone networks and speech codecs," International Telecommunication Union-Telecommunication Standardisation Sector, Tech. Rep., 2001.



**Juan M. Martín Doñas** received the B.Sc. and the M.S. degrees in Telecommunication Engineering from the University of Granada, Spain, in 2015 and 2017, respectively. He is currently working towards the Ph.D. degree on multichannel speech enhancement on mobile devices using deep learning and signal processing techniques. His research interests include speech enhancement, multichannel signal processing and deep learning.



multimedia transmission.

**Ángel M. Gómez** received the M.Sc. and Ph.D. degrees in computing science from the University of Granada, Spain, in 2001 and 2006, respectively. In 2002 he joined the Signal Theory, Networking, and Communications Department, University of Granada, where he is a member of the research group on Signal Processing, Multimedia Transmission and Speech/Audio Technologies (SigMAT). Currently he is an associate professor at the University of Granada. His main interests are on robust speech recognition, speech and audio coding, and



multimedia transmission.

**Antonio M. Peinado** received the M.S. and Ph.D. degrees in physics from the University of Granada, Granada, Spain, in 1987 and 1994, respectively. Since 1988, he has been with the University of Granada, where he has led several research projects related to signal processing and transmission. In 1989, he was a Consultant with the Speech Research Department, AT&T Bell Labs, Murray Hill, NJ, USA. He has held the positions of an Associate Professor from 1996 to 2010 and has been a Full Professor since 2010 with the Department of Signal Theory, Networking and Communications, University of Granada, where he is currently the Head of the research group on Signal Processing, Multimedia Transmission and Speech/Audio Technologies. He has authored numerous publications and has co-authored the book entitled *Speech Recognition Over Digital Channels* (New York, NY, USA: Wiley, 2006). His current research interests are focused on robust speech recognition and transmission, robust image/video transmission, and ultrasound and proteomic signal processing. He has served as a reviewer for several international journals, conferences, and project proposals.

# DESARROLLO DE UN ARRAY AD-HOC DE MICRÓFONOS BASADO EN SMARTPHONES TIPO ANDROID

Tutor: Antonio M Peinado Herreros, Ángel M Gómez García; e-mail: [amp@ugr.es](mailto:amp@ugr.es), [amgg@ugr.es](mailto:amgg@ugr.es)

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

Autor: Pablo Jesús García Martínez, e-mail: [pgarciamartinez@correo.ugr.es](mailto:pgarciamartinez@correo.ugr.es)

**Resumen**—Debido al gran desarrollo tecnológico de los smartphones actuales y también al gran número de estos dispositivos que utilizan el sistema operativo Android, a la vez que su bajo coste, se propone en el siguiente desarrollo la creación de una aplicación Android para la calibración de un array de micrófonos formado por los micrófonos de varios de estos dispositivos, con orientación a implementar, en una segunda fase alguna aplicación de beamforming o, especialmente, tracking con este conjunto de micrófonos.

**Palabras clave**—array, micrófono, calibración, ToA, dToA, tracking, beamforming, Android, Java, Matlab, servidor, aplicación, activity, intent, socket, puerto, clase, dato miembro, método..

## I. INTRODUCCIÓN

Un *array* de micrófonos es un grupo de micrófonos que operan de forma conjunta, de manera que a cada micrófono el sonido le llegará en instantes diferentes, pudiendo determinar de dónde viene éste sonido y hasta cual es la posición de la fuente emisora de señales sonoras.

En este trabajo, más que a array de micrófonos se va a hacer referencia, podríamos decir, a un subtipo de array denominados *WASNs* (*Wireless Acoustic Sensor Networks*). Estos sistemas hacen referencia a un tipo de configuración en la que cada uno de los micrófonos es independiente del otro y tiene capacidades de comunicación inalámbrica, como ocurre con los micrófonos integrados en los smartphones. Además, a diferencia de los arrays de micrófonos tradicionales, que muestrean el sonido solo en un punto a menudo a largas distancias de la fuente, las *WASNs* permiten el uso de un número mayor de micrófonos, a la vez que mejor distribuidos, de manera que se pueden muestrear de una manera más uniforme todo el espacio por el que se propaga la señal sonora, como se puede ver en la figura 1

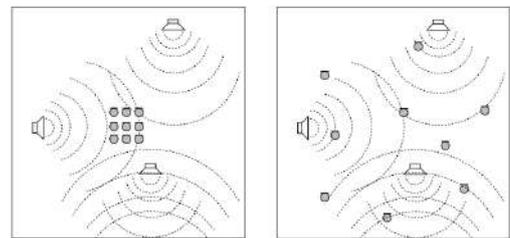


Fig 1: Detalle de un array de micrófonos (izquierda) y una red WASN (derecha)

Estas redes solventan algunos de los principales problemas de los arrays de micrófonos como son la baja SNR de la señal capturada debido a la lejanía del *array* respecto a la fuente sonora y limitaciones de espacio y potencia de cálculo, ya que poseen funcionalidades de procesamiento y de comunicaciones avanzadas. Además es posible establecer diferentes topologías de red, con los mismos dispositivos

Por otra parte necesitamos para el desarrollo posterior de alguna implementación con estas redes, de una calibración de la misma así como de un procesamiento de señal mucho más complejo que para un array convencional.

## II. APLICACIONES QUE MEJORAN USANDO WASNs

**-Hearing Aids (Ayuda a la audición):** Para la reducción del ruido de fondo en ambientes ruidosos y proporcionar una mayor inteligibilidad de la señal de voz. Los micrófonos colocados en los aparatos de audición, uno en cada oído, forman una red de sensores acústicos. Además se podrán conectar a estos aparatos micrófonos inalámbricos externos, de manera de se puede realizar una mejor cancelación cuantos más micrófonos se tengan en la red.

**-Hands-free telephony (Teléfonos manos libres):** El uso de canceladores de ruido en manos libres es habitual en la actualidad. Conociendo el hecho de que cuantos más micrófonos haya en una red, más eficiente será la cancelación, podríamos pensar por ejemplo en un ambiente ruidoso como en un aeropuerto. Si todos los dispositivos presentes en las inmediaciones del aeropuerto formaran una

red, los resultados de la cancelación del ruido permitirían a la gente moverse a lo largo del aeropuerto sin escuchar o transmitir el ruido en sus conversaciones telefónicas.

**-Acoustic Monitoring (Monitorización Acústica).** Los micrófonos tienen la ventaja de que son mucho más baratos que las cámaras y además no tienen el inconveniente de la línea de visión, por tanto se podrían utilizar estos micrófonos para temas relacionados con vigilancia y seguridad.

**-Ambient intelligence (Inteligencia Ambiental):** Para la realización de sistemas inteligentes, por ejemplo de domótica, se pueden utilizar este tipo de redes para captar la voz en donde quiera que esté situado el usuario. cursiva y con un espaciado anterior y posterior de 6 y 3 puntos respectivamente.

### III. APLICACIONES ESPECÍFICAS A LAS QUE SE ORIENTA ESTE TRABAJO

**-Beamforming:** El Beamforming es una técnica de procesamiento de señal que consiste en aumentar la sensibilidad de la respuesta del array en una determinada dirección del espacio. De esta manera se puede procesar la señal deseada captada por la red de sensores e ignorar o mitigar el ruido indeseado que provenga de otra dirección del espacio.

**-Tracking (Localización de Fuentes sonoras):** En escenarios como son las conferencias y exposiciones, es necesario focalizar el audio en los participantes que van participando en la conversación. Para ello puede ser necesario calcular las coordenadas de la fuente sonora, con respecto a las coordenadas de los dispositivos del array.

Para llevar a cabo dichas implementaciones es necesario la calibración del array. Es decir, el cálculo de las coordenadas cartesianas de cada uno de los micrófonos que conforman la red de sensores o *array*.

### IV. SMARTPHONES Y ANDROID

Según el informe anual de ditrendia [1], correspondiente a 2016, a finales de 2015, la penetración de teléfonos móviles ascendió en el mundo al 97%, lo que quiere decir que se añadieron 563 millones durante 2015, subiendo el número total de estos dispositivos a los 7,9 miles de millones de unidades (más que personas en el mundo).

Además, las ventas de smartphones aumentaron en un 13% con respecto al año anterior, un crecimiento principalmente debido a los mercados emergentes de Asia, Pacífico, Japón, Oriente Próximo y África, aunque este crecimiento, si bien seguirá en los próximos años, tenderá a ser más moderado. Según las cifras, si en 2016 hay 4,61 mil millones de usuarios de móviles en el mundo, de los que más de 2,08 mil millones son usuarios de smartphones, en 2019 esta cifra superará los 5,07 mil millones (de los que 2,65 serán smartphones).

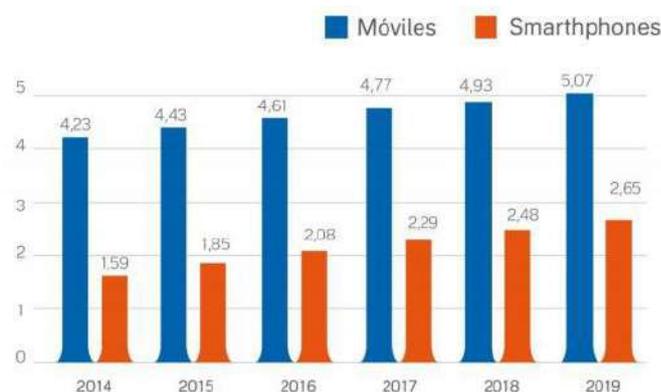


Fig 2: Evolución del número de móviles vs smartphones en el mundo

Por esta razón se ha decidido en el siguiente trabajo utilizar smartphone, por su gran incursión en la sociedad actual por una parte y por su gran versatilidad (capacidades de comunicaciones inalámbricas, grabación de audio, reproducción de audio etc) .

Como sistema operativo para dichos smartphones se ha utilizado Android ya que según IDC [2] Android, es el sistema operativo que llevan alrededor del 85% de los teléfonos inteligentes vendidos el primer cuatrimestre de 2017 y durante los últimos tres años, superando con creces a otros sistemas operativos del mismo nivel, como se puede ver en la Figura 3.

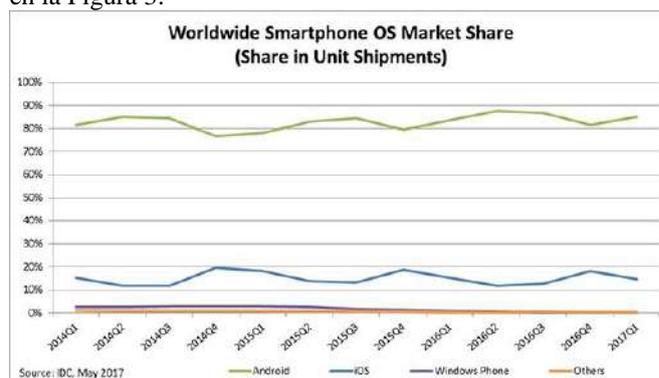


Figura 3. Ocupación de SO en los móviles en el mundo . Verde (Android) Azul (IOs) y Rojo (Otros).

Del mismo modo Android es un SO de libre uso, de manera que puede llegar de forma gratuita a todos los usuarios y ofrece unas grandes capacidades de aprendizaje de uso de su plataforma y de ayuda técnica para desarrolladores.

La versión escogida para la implementación de la aplicación ha sido la versión Android 4.0 Ice Cream Sandwich, ya que las funcionalidades que aporta esta versión son suficientes para el desarrollo de nuestra aplicación y además, a fecha de Junio de 2017 el 90% de los dispositivos Android, utiliza dicha versión, como se puede observar en la Figura 4.

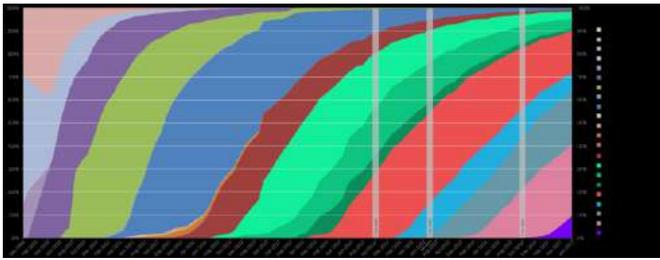


Figura 4: Versiones disponibles de Android según fecha

### V. ESTRUCTURA DEL PROYECTO

Para el desarrollo del siguiente proyecto se van a diferenciar tres bloques con distintas funcionalidades.

**Aplicación Android:** Mediante esta aplicación el usuario debe introducir algunos parámetros necesarios para la correcta deducción de las posiciones de los micrófonos. El objetivo es que estos datos se vayan introduciendo en los  $N$  dispositivos del array sucesivamente para, a continuación, sincronizar a través de los pitidos el array. El usuario es el encargado de hacer correctamente estas tareas.

**Módulo cliente servidor:** Este módulo será el “intermediario” entre la aplicación y la macro que ejecutará el algoritmo deseado. Será una aplicación escrita en Java que va a estar escuchando las peticiones de conexión de cada uno de los dispositivos que forman el array en un puerto, y para cada una de las peticiones abrirá un socket para poder tener la comunicación con el dispositivo. Mediante este socket el servidor tendrá el control del dispositivo, indicándole las acciones a realizar y recibiendo de este los datos necesarios. Se diseñará a su vez un protocolo de comunicaciones entre cliente y servidor que se ajuste a nuestras necesidades. Este protocolo se explicará en detalle posteriormente. Por último este servidor llamará a la macro de Matlab necesaria para ejecutar el algoritmo.

**Script escrito en Matlab:** Este “bloque” es el encargado de proporcionar los resultados finales, es decir, de calcular las posiciones en el espacio euclídeo de cada uno de los micrófonos. Esta macro será llamada desde el servidor cuando estén disponibles todos los datos necesarios para el cálculo.

### VI. EL ALGORITMO DE CALIBRACIÓN

El algoritmo de calibración utilizado para el cálculo de las coordenadas consta de dos sub-algoritmos. Uno de ellos tiene la función de, a través de las grabaciones de las señales de calibración de cada dispositivo, obtener unas coordenadas iniciales. El segundo tiene la función de optimizar estas primeras coordenadas para ofrecer un resultado más ajustado a la realidad.

El algoritmo utilizado para el cálculo de las coordenadas iniciales se denomina algoritmo MDS (Multidimensional Scaling). Dada una matriz de distancias entre los micrófonos de la red (obtenida a partir de los tiempos de llegada de las señales (ToA)),

$$d_{ij} = \|m_i - m_j\| \quad (1)$$

Mediante un proceso de descomposición en valores propios de dicha matriz podemos obtener unas coordenadas iniciales del siguiente modo:

Se puede definir  $B$  en donde  $M$  es la matriz de coordenadas. Una estimación de  $B$  es:

$$\hat{B} = M^{-1} \quad (2)$$

En donde  $\mathbf{1}$  es un vector de longitud  $M$  de unos, e  $I$  es la matriz identidad de orden  $M$ .  $\hat{B}$ , por ser simétrica y semidefinida positiva se puede descomponer como:

$$\hat{B} = U \Lambda U^T \quad (3)$$

Y por último obtener unas coordenadas iniciales quedándonos con los  $p$  (dimensión utilizada, en nuestro caso  $p=2$ ) autovalores y autovectores de mayor valor como:

$$\hat{M} = U_p \Lambda_p U_p^T \quad (4)$$

Una vez obtenidas dichas coordenadas es necesario optimizarlas. Para ello utilizamos el algoritmo dToA. Si antes nos habíamos basado en el tiempo de llegada de la señal desde el altavoz  $k$  al micrófono  $i$ , ahora vamos a utilizar la diferencia de tiempos de llegada dToA, definida como

$$d_{ik} = \|s_i - s_k\| \quad (5)$$

En donde  $s_k$  es la posición del micrófono  $k$  definida como

$$s_k = (m_k \cos \varphi_k, m_k \sin \varphi_k)^T \quad (6)$$

siendo  $\varphi_i$  el ángulo entre una dirección determinada y la recta imaginaria que une el micrófono y el altavoz como se muestra en la figura 5.

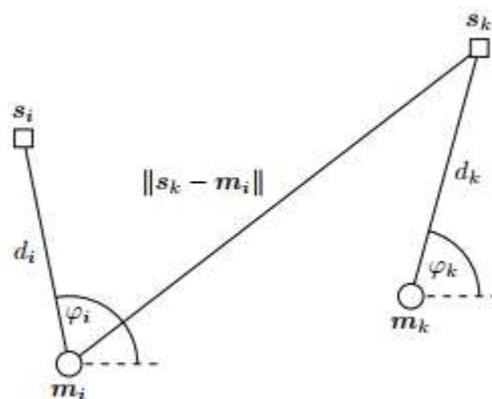


Figura 5: Distancias, orientaciones y configuración espacial del array

y  $t$  son los tiempos de comienzo de la grabación de cada uno de los dispositivos, que es un parámetro también a estimar.

Las medidas tomadas de dToA están afectadas por ruido. Considerando a éste como gaussiano:

$$\hat{t} = \begin{pmatrix} t_1, (\sigma_1) \\ \vdots \\ t_n, (\sigma_n) \end{pmatrix} \quad (7)$$

y por último las coordenadas y tiempos de comienzo los obtenemos de minimizar la siguiente expresión.

$$(\hat{M}, \hat{T}_c) = \left( \frac{(\hat{\tau}_{ij})}{\dots} \right) \quad (8)$$

Con las siguientes restricciones

$$(0, \dots), \quad (9)$$

$$(m, \dots), \quad (10)$$

$$(m, \dots), \quad (11)$$

### VII. EL SERVIDOR JAVA

El servidor Java es una aplicación escrita en el lenguaje de programación Java con la ayuda de NetBeans. Su función principal es la comunicación con los smartphones de la red y por tanto utiliza los *sockets* para realizar ésta función.

Dicha aplicación, que gobierna los dispositivos del array, escucha las peticiones de estos y recibe por último los parámetros necesarios para la obtención de las coordenadas de cada dispositivo. La comunicación se realiza utilizando el modelo cliente (smartphone) servidor (ordenador).

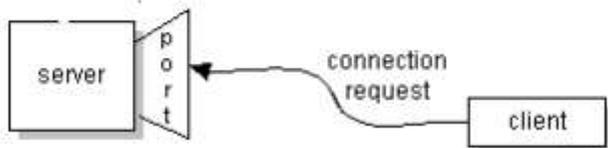


Figura. 6: Petición de servicio del cliente al servidor

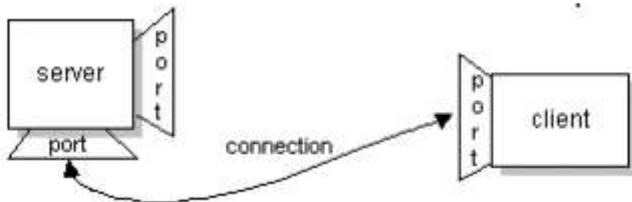


Figura. 7: El servidor da conexión a un cliente

Dicho programa se compone de una única clase, con varios métodos para la estructuración de las funciones y para ejecutarlo deberemos llamarlo desde la línea de comandos.

### VIII. LA APLICACIÓN ANDROID

La aplicación desarrollada es una sencilla aplicación que permite la interacción del usuario del dispositivo con el servidor, además de proporcionar todos los elementos y funcionalidades para la calibración del array que es lo que se desea en este trabajo.

Básicamente, la aplicación utiliza cuatro clases, tres de ellas *Activities*.

La primera actividad se muestra información al usuario sobre lo que deben hacer al abrir la aplicación. Un bloque en el que el usuario debe introducir el valor de la distancia entre altavoz y micrófono del teléfono y, una lista desplegable que contiene valores de ángulos en grados, para ajustar el ángulo formado por el altavoz y la línea central del teléfono como se puede apreciar en la figura 7. Por último un botón, hace que pasemos a la siguiente actividad. Tiene el siguiente aspecto

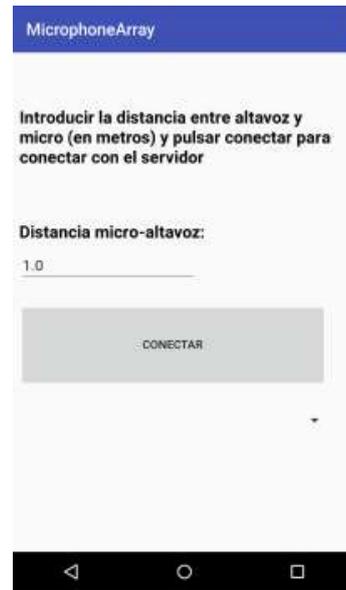


Figura 6: Imagen de la primera actividad de la aplicación

La segunda actividad es una pantalla en blanco, no aporta información alguna al usuario, pero está desarrollando un trabajo en *background*. Es la encargada de abrir los sockets y hacer las peticiones al cliente, para enviar los datos introducidos en la primera actividad. Del mismo modo accede a los sensores internos del dispositivo para recabar información que tendrá que enviar posteriormente al servidor, como el ángulo entre el norte magnético y la recta principal del smartphone (ver figura 7 Eje Y).



Figura. 7: Ángulos a obtener en cada dispositivo

La tercera actividad se compone de un botón con el nombre de sincronización. Cuando este botón ha sido pulsado por todos los dispositivos del array, se comienza la calibración. Cada dispositivo comienza a grabar y a emitir las señales de calibración. Una vez que ha acabado la grabación, se vuelven

a abrir de nuevo sockets para enviar al servidor la grabación y que se pueda ejecutar el algoritmo.

### IX. LAS SEÑALES DE CALIBRACIÓN

se ha escogido como señal de calibración una señal con chirp. Este tipo de señales, hace un barrido frecuencial en el tiempo, es decir a la vez que va aumentando el tiempo aumenta la frecuencia de la señal emitida. Este tipo de señales tiene una gran resolución frecuencial y para obtener unos buenos resultados no es necesario una larga duración de la señal para unos buenos resultados. Algunas pruebas muestran que con una duración aproximada de 100ms se obtienen buenos resultados.

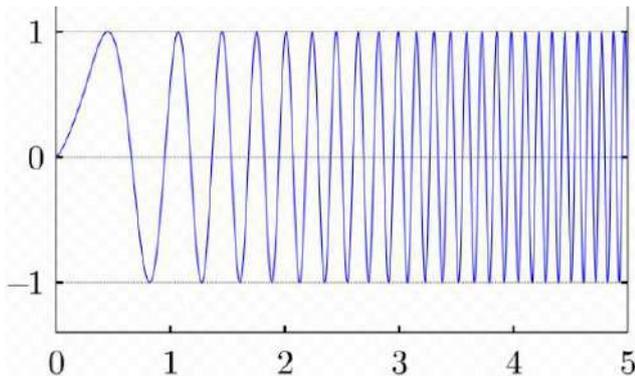


Figura 8: Detalle de una señal con Chirp

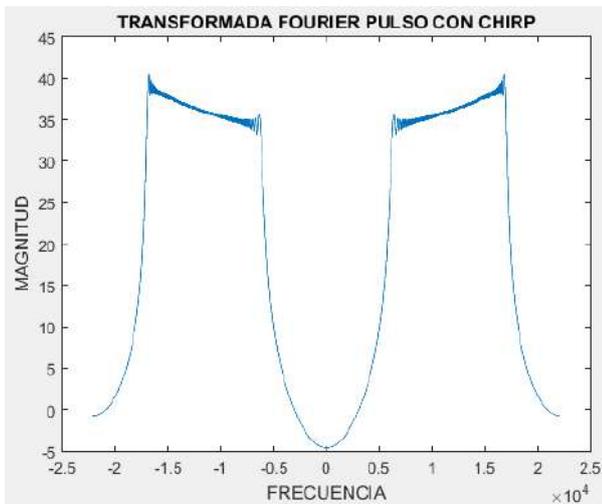


Figura 9: Transformada de Fourier de la señal de chirp empleada

### X. EL PROTOCOLO DE COMUNICACIONES

El protocolo establecido para la correcta operación del sistema se puede ver en la figura 8.

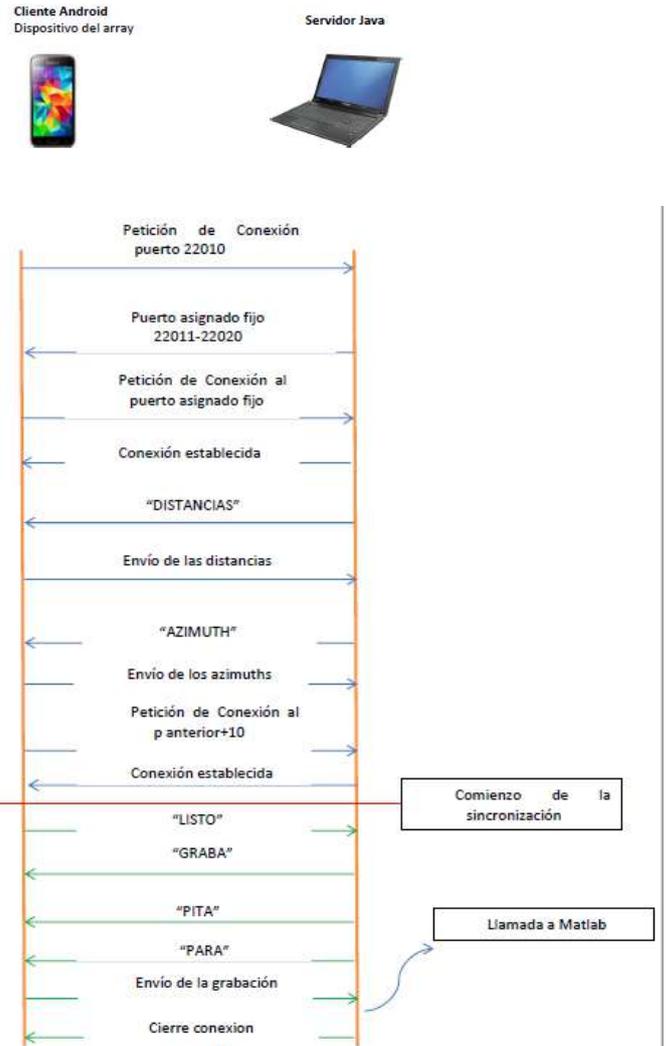


Figura 10: Comunicación smartphone- servidor

Básicamente, cuando se llama al programa desde la línea de comandos para ejecutarlo, se debe de indicar el número de dispositivos que van a formar parte del array, y se creará un objeto ServerSocket que recibirá peticiones de conexión en el puerto 22010. Cuando una petición de conexión le llegue, el servidor rápidamente buscará un puerto libre, en el rango 22011-22020, y por el socket asignado al puerto 22010, transmitirá al cliente el puerto al que debe conectarse. A continuación se abrirá otro objeto ServerSocket (ServerFinal1), el cliente realizará una petición de conexión a este nuevo puerto, para ahora sí, mantener una conexión fija en este nuevo puerto.

Una vez que se ha establecido la conexión en el puerto, el servidor espera que los clientes le manden, por orden las distancias micrófono-altavoz. Para ello el servidor envía el string "DISTANCIAS" y los clientes, al recibirla, envían al servidor dicha distancia que previamente ha sido introducida por el usuario de la aplicación Android. Este procedimiento se hace mediante el método leeDistancias. Las distancias leídas son almacenadas por orden (primero la del dispositivo 1, luego la del 2 etc) en un fichero de texto que se creará en un directorio escogido de nuestro ordenador donde se esté ejecutando el servidor Java, con el nombre distancias.txt.

Del mismo modo y manteniendo la conexión anterior, cuando se han acabado de recibir y guardar correctamente las distancias, el servidor envía a los clientes, de nuevo por orden, el string “AZIMUTH” con el objetivo de que le envíen los azimuths leídos por los sensores de los smartphones que componen el array. Cuando los recibe, de igual modo que para las distancias, los guarda en un fichero azimuth.txt para que posteriormente sea leído por el script de Matlab.

A continuación se debe proceder a la sincronización de los dispositivos del array. Para ello es necesario que los teléfonos se pongan a grabar el audio a su alrededor y que vayan emitiendo señales sonoras (chirps) alternativamente, que van a quedar registradas en la grabación y que posteriormente van a ser determinantes para determinar las coordenadas.

En primer lugar, debido a que en la aplicación en Android se cambia de Activity (se explicará posteriormente), es necesario cerrar todas las conexiones existentes hasta ahora y comenzar unas nuevas en otro puerto. Por facilidad y sencillez, y en pro de evitar conflictos se ha decidido que cada dispositivo se conecte a un puerto nuevo, que se calcula como puerto antiguo+10, es decir, si por ejemplo un dispositivo estaba conectado al puerto 22011 del servidor, ahora automáticamente se conectará al puerto 22021. Esta nueva conexión se hace de manera similar a la anterior, creando un objeto ServerSocket para cada uno de los dispositivos existentes, y esperando hasta que se conecten a este puerto.

Una vez exitosa esta segunda conexión, se hace uso del método mandaGrabar. Cuando los clientes han acabado de hacer las tareas explicadas anteriormente envían el string “LISTO” al servidor. Una vez que el servidor ha recibido dicha cadena de los N dispositivos que componen nuestro array, indica a los dispositivos que pueden iniciar la grabación de audio. Para ello manda la cadena “GRABA”, por orden a cada teléfono.

Una vez que todos los dispositivos están grabando audio, es necesario que se reproduzca el sonido de calibración. Como se ha explicado, anteriormente este sonido es un pulso con chirp de duración variable (se ajustará una duración que aporte buenos resultados). Cada teléfono reproducirá un pitido cuando el servidor le mande la cadena “PITA”. Para no solapar las señales de dos teléfonos, se espera un tiempo prudencial entre dos órdenes de pitido. De toda esta tarea, se encarga el método mandaReproducir.

Una vez reproducidas estas señales sonoras por cada uno de los smartphones, el método mandaParar se encarga de mandar a los clientes el string “PARA” para que detengan la grabación y se inicie el proceso de envío de las grabaciones al servidor.

El servidor entonces, a través del método recibeGrabaciones recibe las grabaciones por orden de cada uno de los dispositivos, la etiqueta en orden y la guarda en el mismo directorio que los ficheros anteriores con formato.raw, y lo deja todo listo para la hora de la ejecución del algoritmo de Matlab

## AGRADECIMIENTOS

En primer lugar me gustaría agradecer a Antonio y Ángel, mis tutores el gran apoyo e interés mostrado en este proyecto, así como la ayuda dada en todos los momentos en los que la he solicitado.

Por otra parte, agradecer a todas las personas que me han ayudado a nivel personal, amigos, familiares, padres etc. Habeis sido mi apoyo fundamental durante todo este tiempo.

Por último, agradecer a todos los profesores y formadores que han enseñado con ahínco todos los conocimientos y que han sido claves para mi formación.

## REFERENCIAS

- [1] Informe ditrendia (Digital Marketing Trends): Mobile en España y en el Mundo 2016.
- [2] Informe Statista: Tasa de penetración de los sistemas operativos para smartphone en España en 2014 y 2015.
- [3] BERTRAND, A. 2011. Applications and trends in wireless acoustic sensor networks: A signal processing perspective. In 2011 18th IEEE Symposium on Communications and Vehicular Technology in the Benelux, SCVT 2011, Anonymous .
- [4] BIRCHFIELD, S.T. 2003. Geometric microphone array calibration by multidimensional scaling. In ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, Anonymous , 157-160.
- [5] HENNECKE, M.H. AND FINK, G.A. 2011. Towards acoustic self-localization of ad hoc smartphone arrays. In 2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays, HSCMA'11, Anonymous , 127-132.
- [6] PLINGE, A., JACOB, F., HAEB-UMBACH, R. AND FINK, G.A. 2016. Acoustic Microphone Geometry Calibration: An overview and experimental evaluation of state-of-the-art algorithms. IEEE Signal Processing Magazine 33, 14-29.
- [7] JANSON, T., SCHINDELHAUER, C. AND WENDEBERG, J. 2010. Self-localization application for iPhone using only ambient sound signals. In 2010 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2010 - Conference Proceedings, Anonymous .
- [8] TAGHIZADEH, M.J., PARHIZKAR, R., GARNER, P.N., BOURLARD, H. AND ASAEI, A. 2015. Ad hoc microphone array calibration: Euclidean distance matrix completion algorithm and theoretical guarantees. Signal Processing 107, 123-140.
- [9] ASAEI, A., MOHAMMADIHA, N., TAGHIZADEH, M.J., DOCLIO, S. AND BOURLARD, H. 2015. On application of non-negative matrix factorization for ad hoc microphone array calibration from incomplete noisy distances. In ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, Anonymous , 2694-2698.
- [10] HIMAWAN, I. 2010. Speech recognition using ad-hoc microphone arrays.
- [11] V. Rabinkin, Daniel & J. Renomeron, Richard & Dahl, Arthur & French, Joseph & L. Flanagan, James & H. Bianchi, Michael. (1997). A DSP Implementation of Source Location Using Microphone Arrays. The Journal of the Acoustical Society of America. 99. . 10.1121/1.415678.
- [12] NetBeans: <https://netbeans.org>.
- [13] Ming Xue y Changjun Zhu, Changchun Institute of Technology, “The Socket Programming and Software Design for Communication Based on Client/Server”, 2009.
- [14] Programación Android y Android Studio : <https://developer.android.com>.
- [15] <https://es.mathworks.com>.
- [16] Talantzis, F & Constantinides, A. (2007). The effect of microphone array geometry on acoustic source localization. 7. 4508-4517.
- [17] Gironés, J. T. (2012). El gran libro de Android. Marcombo.
- [18] Sznajdleder, P. (2011). Java a fondo. Estudio del lenguaje y desarrollo de aplicaciones. Argentina: RA-MA.
- [19] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero III, R. L. Moses, and N. S. Correal, “Locating the nodes: cooperative localization in wireless sensor networks,” vol. 22, no. 4, pp. 54–69, 2005
- [20] Z. Liu, Z. Zhang, L.-W. He, and P. Chou, “Energy-Based Sound Source Localization and Gain Normalization for Ad Hoc Microphone Arrays,” in IEEE Int. Conf. Acoust., Speech, Signal Process., 2007, pp. II-761–II-764.

# DESARROLLO DE UN SISTEMA DE BEAMFORMING BASADO EN REDES NEURONALES PROFUNDAS

Autor: David Pareja Martos, e-mail: dpm92@correo.ugr.es

Tutor: Ángel Manuel Gómez García, e-mail: amgg@ugr.es

Tutor: Antonio Miguel Peinado Herreros, e-mail: amp@ugr.es

Titulación: Máster Universitario en Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones

Universidad de Granada

**Resumen**—En este trabajo se presenta un sistema para combatir el ruido a la hora de reconocer la información transmitida por la voz humana. Capturando la voz mediante un dispositivo con múltiples micrófonos, puede hacerse uso de la información espacial de dichos micrófonos para mitigar el ruido mediante un sistema de beamforming y mejorar la calidad de voz. Para caracterizar el sistema de beamformer, se va a entrenar una red neuronal profunda (*Deep Neural Network*, DNN) mediante una base de datos de locuciones. Una vez entrenada, la DNN se encargará de estimar los parámetros necesarios para que el beamformer mejore las señales de forma efectiva.

**Palabras clave**—Estimación online, máscaras espectrales, MVDR, redes neuronales profundas.

## I. INTRODUCCIÓN

LAS comunicaciones de voz, ya sea entre personas o interactuando con una máquina, siempre han tenido un enemigo tradicional: el ruido, entendiendo como ruido en este caso toda señal que pueda perjudicar al entendimiento de la información que se desea transmitir. Este ruido afecta, entre otras cosas, a la inteligibilidad del habla por parte de las personas, y a la dificultad de las máquinas de reconocer la información recibida. Por ello, en ambos casos supone una gran ventaja tener la capacidad de mitigar estos elementos ruidosos.

Los dispositivos móviles modernos como tabletas o smartphones incorporan múltiples micrófonos, lo cual les permite capturar la misma señal sonora en versiones ligeramente diferentes. Estas diferencias permiten llevar a cabo un procesamiento posterior para mejorar la parte de la información interesante y mitigar la que perjudica su reconocimiento. Una de estas técnicas es la conocida como conformación del haz o *beamforming*.

Por otro lado, precisamente la naturaleza ruidosa de la mayoría de escenarios en los que el habla es capturada, dificulta caracterizar la manera en que el ruido afecta a la señal, complicando así el uso de beamforming. En general, el beamformer a utilizar depende en gran medida de las características del canal de transmisión y la distribución geométrica de los micrófonos. Por ejemplo, en [1] se utilizó un array de micrófonos con distribución esférica, en el que se empleó un beamformer *Delay and Sum* (DAS). Para lograr la caracterización del beamformer, en este trabajo se ha hecho uso de DNNs.

En los últimos años, el uso conjunto de DNNs y beamforming ha cobrado un gran auge en popularidad debido a su efectividad, habiendo un gran número de trabajos al respecto. Por ejemplo, en [2] se propuso un sistema para reconocimiento de habla distante con un array de micrófonos lejanos, usando una red neuronal convolucional y un DAS como beamformer. Otra aplicación ha sido la estimación de máscaras de probabilidad de presencia de voz (*Speech Presence Probability*, SPP) y de ruido, usando una *Bidirectional Long Short-Term Memory* (BLSTM) y *Minimum Variance Distortionless Response* (MVDR) y *Generalized Eigenvalue* (GEV) como beamformers en [3]. Si bien en el anterior se busca una mejora en el reconocimiento automático del habla, en el presente se trata solamente el realce de voz. El MVDR es un beamformer que se basa en minimizar las componentes ruidosas sin distorsionar la señal de voz. Las redes LSTM se caracterizan por tener la capacidad usar la información del pasado en el procesamiento de la información del presente. Sus versiones bidireccionales, las BLSTM, adicionalmente tienen en cuenta la información del futuro. Estos mismos beamformers (MVDR y GEV) y la red BLSTM fueron utilizados en [4] y [5], pero con un esquema distinto de extracción de los vectores de características y de postprocesado.

En trabajos más recientes se ha llevado a cabo estimación robusta de máscaras en [6], también mediante uso de BLSTM y el beamformer GEV. En [7] se emplea un beamformer basado en estimación iterativa de máscaras junto con una sucesión de capas de red neuronal hacia delante (*Feed-Forward*, FF) y BLSTM conectadas paralelamente. El objetivo de ambos trabajos es estimar máscaras de elementos tiempo-frecuencia que representan la probabilidad de presencia de voz o de ruido en la transformada de Fourier de tiempo corto (*Short Time Fourier Transform*, STFT) de una señal sonora.

En el presente trabajo se propone una estructura similar a la de [3]. Se emplea un beamformer MVDR cuyos parámetros se van a obtener mediante máscaras de SPP y de probabilidad de ruido que se van a estimar con la DNN. A diferencia de [3], se va a emplear una capa LSTM en lugar de una BLSTM para poder utilizar el sistema de forma online.

La estructura del resto del documento es la descrita a continuación. En la sección II se expone la fórmula matemática del MVDR y se desarrollan todos los parámetros a estimar para su obtención. En la sección III se detalla la imple-

mentación del sistema, indicando paso a paso el recorrido de las señales al atravesarlo. En la sección IV se desglosa el marco de trabajo, describiendo la base de datos utilizada, las métricas empleadas para los resultados, y los parámetros utilizados para computar la STFT, así como el diseño concreto de la DNN. En la sección V se presentan los resultados obtenidos. Por último, en la sección VI se exponen las conclusiones extraídas del trabajo y sus resultados.

## II. ESTIMACIÓN DE PARÁMETROS MEDIANTE MÁSCARAS

Con el fin de lograr la mitigación de ruido, en este trabajo se ha empleado un beamformer MVDR, cuya obtención se logra de la siguiente forma,

$$\mathbf{w} = \frac{[\Phi_{VV} + \mu I]^{-1} \mathbf{d}_s}{\mathbf{d}_s^H [\Phi_{VV} + \mu I]^{-1} \mathbf{d}_s} \quad (1)$$

donde  $\mathbf{w}$  son los filtros del beamformer,  $\Phi_{VV}$  es la matriz de correlaciones espaciales (*Spatial Correlation Matrix*, SCM) del ruido,  $\mathbf{d}_s$  es el *steering vector* y  $\mu I$  es un escalar muy pequeño multiplicado por la matriz identidad, que se usa para asegurar que la SCM del ruido sea invertible.

### A. Máscaras de probabilidad

Para poder obtener la SCM del ruido y el steering vector es necesario un estimador, ya que en escenarios reales no son conocidos. La técnica empleada en este trabajo para estimarlos es el uso de máscaras para representar la SPP y la probabilidad de ruido. Estas máscaras constan de un valor entre 0 y 1 para cada bin tiempo-frecuencia de la STFT de las señales de voz y de ruido, indicando la probabilidad de presencia de voz y de ruido, respectivamente. La DNN se encargará de estimar estas máscaras, mientras que para obtener los targets necesarios para entrenar la red se va a hacer uso de máscaras binarias ideales (*Ideal Binary Mask*, IBM), que serán detalladas en la sección de implementación.

### B. Matriz de correlaciones espaciales

Utilizando las máscaras estimadas de voz y de ruido, se pueden estimar a su vez las SCMs de voz y de ruido. Para ello, se presentan dos métodos, diferenciados por el tipo de sistema que suponen.

**Estimación offline:** requiere disponer de toda la información de una secuencia dada para poder ser procesada. En este caso se propone la utilizada en [3], que tiene la siguiente forma, donde se omite la dependencia con la frecuencia para facilitar la lectura, asumiendo una frecuencia dada,

$$\Phi_{XX} = \sum_{t=1}^T M_X \mathbf{Y}(t) \mathbf{Y}(t)^H \quad (2)$$

$$\Phi_{VV} = \sum_{t=1}^T M_V \mathbf{Y}(t) \mathbf{Y}(t)^H \quad (3)$$

siendo  $\mathbf{Y}(t)$  la STFT de la señal ruidosa,  $\Phi_{XX}$  la SCM de la voz,  $\Phi_{VV}$  la SCM del ruido,  $M_X$  la máscara de probabilidad de voz y  $M_V$  la máscara de probabilidad de ruido.

**Estimación online:** permite ir calculando la SCM recursivamente sin necesidad de disponer de todos los datos a la vez.

Para ello, en este caso se propone el procedimiento utilizado en [14], que tiene la siguiente forma,

$$\hat{\Phi}_{VV}(t, k) = \alpha(t, k) \hat{\Phi}_{VV}(t-1, k) + (1 - \alpha(t, k)) \mathbf{y}(t, k) \mathbf{y}^H(t, k) \quad (4)$$

donde  $\mathbf{y}(t, k)$  es la STFT de la señal ruidosa y  $\alpha(t, k)$  es un parámetro dependiente de la SPP dado por,

$$\alpha(t, k) = \tilde{\alpha} + p_x(t, k)(1 - \tilde{\alpha}) \quad (5)$$

donde  $p_x(t, k)$  es la SPP (en este caso, la máscara de voz  $M_X$ ) y donde  $\tilde{\alpha} \in [0, 1)$  es una constante que determina el rango efectivo de  $\alpha$ ; más concretamente,  $\alpha(t, k) \in [\tilde{\alpha}, 1]$ . Para la estimación de la SCM de voz se sigue el procedimiento análogo, utilizando la máscara de ruido como SPP.

### C. Steering vector

El steering vector se puede estimar mediante la descomposición en autovalores (*EigenValue Descomposition*, EVD) de la SCM de la voz [8], obteniendo lo siguiente,

$$\Phi_{XX} = \mathbf{A} \mathbf{A}^H \Phi_S = \mathbf{v}_{S_1} \mathbf{v}_{S_1}^H \lambda_{S_1} \quad (6)$$

donde  $\mathbf{A}$  es la matriz de funciones de transferencia acústicas (*Acoustic Transfer Function*, ATF),  $\Phi_S$  es la densidad de potencia espectral de voz monocanal, y  $\lambda_{S_1}$  y  $\mathbf{v}_{S_1}$  son el único autovalor y el único autovector de  $\Phi_S$ , ya que esta matriz es de rango 1 para un único locutor. Normalizando la ATF respecto de uno de los canales, se obtiene la función de transferencia relativa (*Relative Transfer Function*, RTF) de la voz, que en este caso actúa como steering vector.

## III. IMPLEMENTACIÓN

### A. Proceso general del sistema

En la figura 1 se expone el proceso general o pipeline del sistema, que abarca desde que se extraen las señales multicanal de la base de datos hasta que se obtiene la señal mejorada. La DNN es previamente entrenada con el conjunto de entrenamiento. Para ello, se han usado versiones de las señales que contienen solo la voz limpia y solo el ruido de fondo. Con estas versiones se generan las IBMs, que consisten en, dado un conjunto de datos que pueden pertenecer a una de dos clasificaciones, asignar un valor binario a cada elemento para indicar si el elemento pertenece a una clasificación o a otra (con un 0 o un 1). Aplicando esta técnica al caso de una señal STFT de voz limpia puede estimarse si en un bin tiempo-frecuencia hay presencia o ausencia de voz, y análogamente puede estimarse la presencia o ausencia de ruido con una señal STFT de sólo ruido. En este caso, se propone la técnica utilizada en [3],

$$\text{IBM}_X(t, f) = \begin{cases} 1, & \frac{\|\mathbf{X}\|}{\|\mathbf{N}\|} > 10^{\text{th}_X(f)}, \\ 0, & \text{en otro caso,} \end{cases}$$

$$\text{IBM}_N(t, f) = \begin{cases} 1, & \frac{\|\mathbf{X}\|}{\|\mathbf{N}\|} < 10^{\text{th}_N(f)}, \\ 0, & \text{en otro caso.} \end{cases}$$

donde respectivamente  $IBM_X$  y  $IBM_N$  son las máscaras binarias de voz y de ruido,  $\mathbf{X}$  y  $\mathbf{N}$  los espectros STFT de voz y de ruido, y  $th_X$  y  $th_N$  los umbrales sobre los que se decide si hay presencia o ausencia de voz o de ruido.

Las IBMs se usarán como targets comparándolas con las máscaras estimadas por la red mediante la función de coste. Esta información del coste recibida en el optimizador será propagada hacia atrás y se actualizarán los parámetros de la red mediante el algoritmo de optimización.

### B. Procesado pre-DNN

En primer lugar, en cada iteración del sistema se leerán tantos ficheros del conjunto escogido de la base de datos como indique el tamaño del batch que se haya definido. De cada uno se obtendrá la señal en el tiempo ruidosa  $y$ . A continuación, se calcula la STFT, dando lugar a la señal  $Y$ . Después de este paso, se calcula el módulo de la señal STFT recién obtenida, dando lugar a la señal  $|Y|$  en el pipeline. Esta señal se normaliza mediante los vectores de medias y de varianzas  $\mu_Y$  y  $\sigma_Y^2$  obtenidos del conjunto de entrenamiento, dando lugar a la señal  $|\tilde{Y}|$ , que actuará de entrada a la red neuronal.

### C. Estimación de máscaras con DNN

La red neuronal ya entrenada recibirá como entrada la señal  $|\tilde{Y}|$ , y proporcionará a la salida las estimaciones de las máscaras correspondientes de voz  $\hat{M}_X$  y de ruido  $\hat{M}_V$ . Con estas máscaras, se puede proceder a calcular los parámetros del beamformer para obtener la señal mejorada.

### D. Estimación de parámetros y reconstrucción de señales

Una vez se tienen las estimaciones de las máscaras multicanal  $\hat{M}_X$  y  $\hat{M}_V$ , se va a computar la mediana de cada una, para tener en cada bin tiempo-frecuencia el valor más predominante de entre los canales. Estas máscaras estimadas pasan a continuación a utilizarse junto con la señal ruidosa multicanal original  $Y$  para estimar las SCMs de voz  $\Phi_{XX}$  y de ruido  $\Phi_{VV}$ . A continuación, se computa la EVD de la SCM de voz para obtener la RTF como steering vector.

Conocidas la SCM de ruido y el steering vector, el siguiente paso es la estimación de los pesos  $\mathbf{w}$  del beamformer MVDR, lo cual se calcula simplemente aplicando la ecuación (1). Disponiendo ya del beamformer, se realiza el filtrado espacial a la señal ruidosa multicanal original  $Y$  y de esta manera se logra la estimación de la señal de voz monocanal  $\hat{X}$  en el dominio de la STFT. Finalmente, se lleva a cabo la STFT inversa, dando lugar a la señal mejorada  $\hat{x}$  en el dominio del tiempo.

## IV. MARCO DE TRABAJO

### A. Base de datos

La base de datos utilizada en este trabajo es la de los desafíos CHiME3 [9] y CHiME4 [10] de separación y reconocimiento de voz, que consta de locuciones grabadas con una tableta de seis micrófonos en distintos ambientes ruidosos. Para este trabajo se ha empleado una parte de ella para entrenamiento y test del sistema. Este fragmento se divide

Tabla I  
ESTRUCTURA DE LA RED NEURONAL IMPLEMENTADA.

	Unidades	Tipo	No linealidad	$p_{dropout}$
L1	513	LSTM	Tanh	0.5
L2	513	FF	ReLU	0.5
L3	513	FF	ReLU	0.5
L4	1026	FF	Sigmoid	0.0

a su vez en tres clasificaciones: Por un lado, hay un conjunto para entrenamiento de la DNN y otro para test (extracción de resultados). Por otro lado, las locuciones han sido grabadas en cuatros localizaciones diferentes: un bus, una cafetería, un área peatonal y una calle. Por último, de cada locución se disponen de versiones con la componente de voz limpia y versiones con la componente de ruido.

### B. Métricas perceptuales

Con el fin de evaluar los resultados alcanzados al utilizar el sistema implementado para mejorar las señales de voz, se van a utilizar dos métricas que miden cómo de perceptualmente buena es una señal respecto a otra de referencia.

1) *PESQ*: La evaluación perceptual de calidad de voz (*Perceptual Evaluation of Speech Quality*, PESQ) otorga un coeficiente de la métrica de opinión media (*Mean Opinion Score*, MOS) comparando la señal degradada con respecto a la señal limpia [11].

2) *STOI*: La medida de inteligibilidad del objetivo en tiempo corto (*Short-Time Objective Intelligibility*, STOI) también otorga una medida de similitud entre señal de referencia y degradada, basándose en un coeficiente de correlación entre las envolventes temporales de dichas señales [12].

### C. Parámetros de la STFT

La operación de STFT de las señales de audio se va a llevar a cabo con  $N = 1024$  puntos, y se va a hacer enventanado mediante una ventana de Hanning de  $w_{size} = 1024$  muestras, con un desplazamiento de la ventana de  $w_{shift} = 256$  muestras, tal y como se mostró en el pipeline del sistema. También van a añadirse ceros en los extremos mediante zero-padding para eliminar los efectos de los bordes y poder hacer reconstrucción perfecta. Al final del pipeline, se lleva a cabo la transformada inversa de la señal estimada con el beamformer, haciendo uso de los mismos parámetros descritos de tamaño y enventanado que se aplicaron con la STFT.

### D. Parámetros de la red neuronal

1) *Capas de la red*: En este trabajo se ha replicado el diseño de DNN utilizado en [3] y [13], siendo la red neuronal utilizada la mostrada en la tabla I. Esta red neuronal es idéntica a la presentada en los trabajos referenciados, con una diferencia: la primera capa en dichos trabajos es una BLSTM, mientras que en este caso es LSTM unidireccional. Esto se ha hecho para permitir que el sistema sea online ya que, de utilizar una capa BLSTM, si bien es una técnica más potente que la LSTM estándar, requiere sentencias completas y por tanto forzaría al sistema ser offline.

La primera capa consta de 513 neuronas y es LSTM, usando como función no lineal la tangente hiperbólica y un dropout

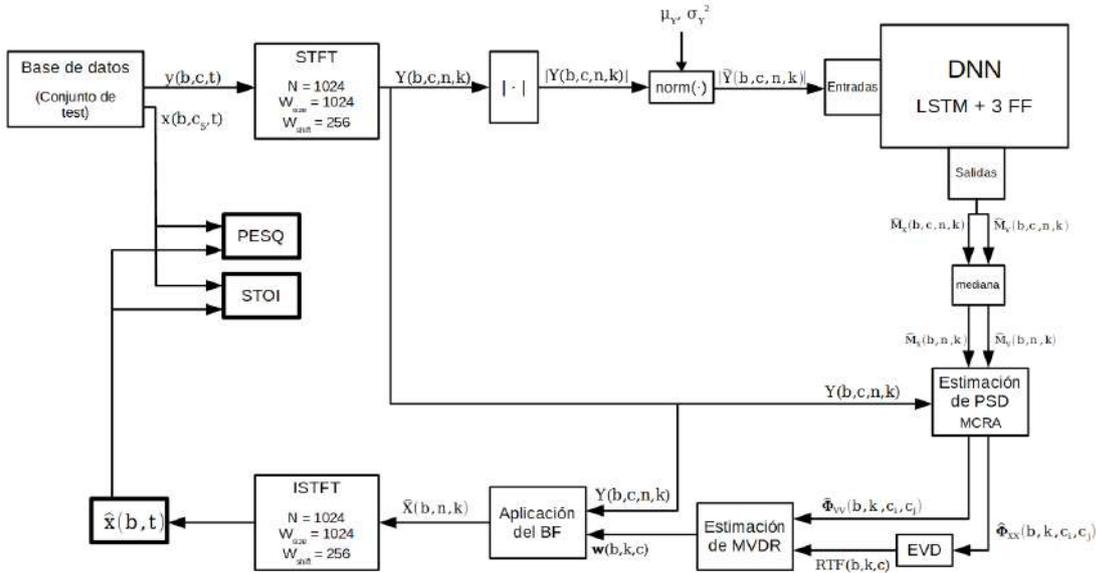


Fig. 1. Pipeline del sistema.

de 0.5. Las capas 2 y 3 son idénticas, de tipo FF con una ReLU como no linealidad y también un dropout de 0.5. Por último, la capa 4 tiene 1026 unidades, también es de tipo FF y tiene una función sigmoide como no linealidad, sin dropout.

La razón de esta estructura tan concreta es que las características que se van a pasar de entrada a la red son las STFT de las señales de audio  $y$ , como ya se ha mencionado, se va a computar con  $N = 1024$  puntos, lo cual hace que toda la información esté en los  $N/2 + 1 = 513$  primeros puntos del espectro, que son el número de unidades de entrada a la red. Por otro lado, la capa de salida tiene 1026 unidades, justo el doble, ya que los parámetros que se van a estimar son las máscaras de voz y de ruido, que requerirán 513 unidades cada una ya que tienen las mismas dimensiones que el espectro de la STFT.

2) *Función de coste*: En este trabajo se ha utilizado la entropía cruzada binaria,

$$l(x, y) = L = \{l_1, \dots, l_N\}^T, \text{ siendo:} \quad (7)$$

$$l_n = -w_n [y_n \cdot \log x_n + (1 - y_n) \cdot \log(1 - x_n)] \quad (8)$$

donde  $N$  es el tamaño del batch,  $x$  son los valores estimados,  $y$  son los targets, y  $w_n$  son pesos a aplicar a cada elemento si se desea. Esta función es idónea para clasificar la pertenencia o no a un conjunto, como es el caso de presencia o ausencia de voz y de ruido en las máscaras.

3) *Optimizador*: En este trabajo se ha escogido la estimación adaptativa del momento (*ADaptive Moment estimation*, ADAM) [15]. Con este algoritmo, el entrenamiento se ha llevado a cabo durante cinco epochs aplicando al optimizador un coeficiente de aprendizaje de  $10^{-4}$ . Se ha ejecutado este número de epochs porque ADAM converge rápidamente y no es deseable sufrir sobreajuste.

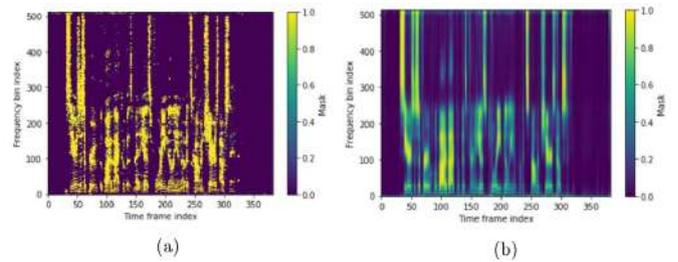


Fig. 2. Máscara de voz oráculo (a) y máscara de voz estimada por la DNN (b).

## V. RESULTADOS

### A. Máscaras estimadas

En la figura 2 se muestra una comparativa de máscaras de voz de un ejemplo del set de test. En la imagen (a) se expone la máscara binaria oráculo conociendo el espectro de la voz. En la imagen (b) se muestra la máscara estimada por la red neuronal. Se observa que hay una gran similitud entre ambas a grandes rasgos, si bien a escala más pequeña se aprecia mucha mayor definición en la máscara oráculo debido a que es binaria y no hay transición de un bin tiempo-frecuencia a otro, mientras que en la máscara estimada se aprecian transiciones suaves ya que los bins tiempo-frecuencia estimados sí contienen valores intermedios entre 0 y 1.

En la figura 3 se muestra una comparativa entre la STFT de la señal ruidosa (a) y la STFT de la señal mejorada con el beamformer (b). En ella se aprecia cómo en cada trama los bins pertenecientes al ruido se ven en general atenuados, reflejándose en un oscurecimiento del fondo del espectrograma. A su vez, la energía de los bins pertenecientes a voz, que son los mostrados en las máscaras de la figura 2, permanecen casi intactos, haciendo que la STFT de la señal mejorada se parezca en forma a la propia máscara de voz.

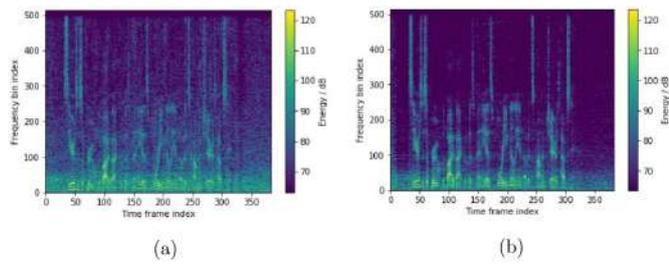


Fig. 3. STFT de señal ruidosa (a) y STFT de señal mejorada (b).

Tabla II  
PUNTUACIONES MEDIAS DE PESQ Y STOI DEL CONJUNTO DE TEST.

		PESQ	STOI
Señal sin procesar		2.169	0.8635
Másc. oráculo	Offline	2.643	0.9640
	MCRA	2.873	0.9663
Másc. DNN	Offline	2.533	0.9413
	MCRA	2.497	0.9035

### B. Métricas perceptuales

A continuación se exponen los resultados obtenidos mediante las métricas descritas en el marco experimental. Estos resultados se han tomado del conjunto de test.

En la tabla II se muestran las puntuaciones medias de PESQ y STOI. Las filas de la tabla tienen varias subdivisiones. En primer lugar, se exponen las puntuaciones medias de las señales sin procesar, y a continuación se exponen las puntuaciones de las señales reconstruidas en función de si se han obtenido con máscaras oráculo o con las máscaras estimadas por la DNN. Dentro de esta última división, se hace la distinción de si la SCM se ha obtenido por el método offline o por el método online MCRA. Por columnas, se exponen a la derecha las puntuaciones de PESQ y STOI.

Se observa que con el sistema offline se obtiene una media de puntuación del PESQ de 2.533 puntos, que supone una mejora media de 0.364 puntos respecto a la señal ruidosa. Con el sistema online, sin embargo, la puntuación es de 2.497, que supone una mejora de 0.328 respecto a la señal ruidosa, y una diferencia respecto al sistema offline de 0.035 que, dependiendo de la aplicación, puede considerarse un pequeño precio a pagar por hacer que el sistema sea capaz de ir procesando la información sin necesidad esperar a haber cada secuencia completa. De manera similar ocurre con el STOI, logrando un índice de inteligibilidad de 0.9413 para la estimación offline, y de 0.9035 para la estimación online, superándose así la marca de 0.9 en ambos casos, cuando las señales ruidosas obtienen una media de 0.8635. Con las máscaras oráculo, como es de esperar, se obtienen puntuaciones más altas en todos los casos, que sirven de cota superior para indicar el posible margen de mejora que tiene el sistema partiendo de las versiones ruidosas. En este aspecto, cabe resaltar que los resultados obtenidos no se separan mucho del máximo alcanzable, especialmente con el STOI en estimación offline.

En la tabla III se exponen las puntuaciones medias en función del tipo de ruido, de manera que se hace la misma subdivisión que en la tabla II pero se añade una columna a la izquierda que indica la localización (bus, café, área peatonal y

Tabla III  
PUNTUACIONES MEDIAS DE PESQ Y STOI POR TIPO DE RUIDO

		PESQ	STOI	
BUS	Señal sin procesar	2.268	0.8683	
	Másc. oráculo	Offline	2.826	0.9777
		MCRA	2.993	0.9701
	Másc. DNN	Offline	<b>2.663</b>	<b>0.9490</b>
MCRA		2.562	0.9052	
CAF	Señal sin procesar	2.005	0.8440	
	Másc. oráculo	Offline	2.435	0.9520
		MCRA	2.703	0.9608
	Másc. DNN	Offline	<b>2.353</b>	<b>0.9306</b>
MCRA		2.336	0.8895	
PED	Señal sin procesar	2.261	0.8867	
	Másc. oráculo	Offline	2.705	0.9695
		MCRA	2.946	0.9728
	Másc. DNN	Offline	<b>2.618</b>	<b>0.9552</b>
MCRA		2.603	0.9234	
STR	Señal sin procesar	2.140	0.8550	
	Másc. oráculo	Offline	2.605	0.9567
		MCRA	2.850	0.9614
	Másc. DNN	Offline	<b>2.500</b>	<b>0.9304</b>
MCRA		2.489	0.8960	

calle) en la que las señales fueron grabadas, de manera que en las columnas de PESQ y STOI se tienen los valores medios de ese tipo de ruido exclusivamente.

Se puede comprobar que, si bien las puntuaciones base de PESQ difieren según el tipo de ruido, la mejora alcanzada con el sistema tanto con la estimación offline como con MCRA ronda una diferencia de aproximadamente 0.35 puntos respecto a las señales sin procesar, si bien la estimación offline otorga puntuaciones ligeramente más altas. La excepción está en la localización del bus, en la que la diferencia de puntuación en PESQ obtenida con el sistema offline es de unos 0.40 puntos mayor que las señales sin procesar, mientras que con MCRA solo se ganan unos 0.30 puntos. Aparte de esta salvedad, el hecho de que la mejora sea similar en los distintos escenarios implica que se ha logrado conseguir cierta generalidad en el entrenamiento de la red. Las puntuaciones de STOI también son consistentes en este sentido. Si bien se logra una menor mejora con el sistema online, se está obteniendo en cada escenario siempre una mejora de en torno a 0.8 puntos con el sistema offline, y unos 0.4 con MCRA.

Comparando los resultados de la tabla II con los de [3], en aquel trabajo lograron, haciendo uso de la red BLSTM, una puntuación media de PESQ para el conjunto de *development* de CHiME solamente 0.1 puntos superior (en torno a 2.6) a la obtenida en este trabajo. Cabe destacar que, además del mayor rendimiento que supone el uso de la BLSTM sobre la LSTM unidireccional, obtuvieron este resultado usando un beamformer GEV, sobre el cual reportaron en el trabajo que les otorgó resultados algo mejores que el MVDR debido a la inversión de matriz del ruido que éste requiere. Por lo tanto, se estima que se han logrado unos resultados exitosos acordes al sistema implementado.

## VI. CONCLUSIONES

A la vista de los resultados, se ha logrado implementar un sistema capaz de mejorar la calidad de una señal multi-canal procesada mediante un beamformer estimado con redes neuronales, logrando generalizarlo para su uso en entornos ruidosos cotidianos. El uso de un sistema online permite

además adaptarlo a aplicaciones que no requieran tener toda la información para ir procesando los datos.

Se ha conseguido obtener unos resultados comparables a los que se lograron en [3] en cuanto a métricas de PESQ, si bien no se han utilizado exactamente los mismos recursos que se hicieron en el mencionado trabajo, al no haber hecho uso de BLSTM.

#### REFERENCIAS

- [1] A. Fahim, P. N. Samarasinghe, T. D. Abhayapala. (2018, may.). PSD estimation and source separation in a noisy reverberant environment using a spherical microphone array. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. vol. 26, págs. 1594-1607.
- [2] P. Swietojanski, A. Ghoshal, S. Renals. (2014, may.). Convolutional Neural Networks for Distant Speech Recognition. *IEEE Signal Processing Letters*. vol. 21, págs. 1120-1124.
- [3] J. Heymann, L. Drude, R. Haeb-Umbach. (2016, mar.). Neural Network Based Spectral Mask Estimation for Acoustic Beamforming. *IEEE International Conference on Acoustics, Speech and Signal Processing*.
- [4] Z. Meng, S. Watanabe, J. R. Hershey, H. Erdogan. (2017, jun.). Deep long short-term memory adaptive beamforming networks for multichannel robust speech recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing*.
- [5] H. Erdogan, J. R. Hershey, S. Watanabe. (2016, sep.). Improved MVDR beamforming using single-channel mask prediction networks. *Inter-speech*.
- [6] Y. Zhou, Y. Qian. (2018, abr.). Robust mask estimation by integrating neural network-based and clustering-based approaches for adaptive acoustic beamforming. *IEEE International Conference on Acoustics, Speech and Signal Processing*.
- [7] Y. Tu, J. Du, L. Sun, F. Ma, H. Wang, J. Chen. (2018, nov.). An iterative mask estimation approach to deep learning based multi-channel speech recognition. *Speech Communication*.
- [8] S. Markovich, S. Gannot, I. Cohen. (2009, jun.). Multichannel Eigenspace Beamforming in a Reverberant Noisy Environment With Multiple Interfering Speech Signals. *IEEE Transactions on Audio, Speech, and Language Processing*. vol. 17, págs. 1071-1086.
- [9] The 3rd CHiME Speech Separation and Recognition Challenge. (2015).
- [10] The 4th CHiME Speech Separation and Recognition Challenge. (2016). Disponible: [http://spandh.dcs.shef.ac.uk/chime\\_challenge/chime2016/](http://spandh.dcs.shef.ac.uk/chime_challenge/chime2016/)
- [11] A. W. Rix, J. G. Beerends, M. P. Hollier, A. P. Hekstra. (2001, may.). Perceptual Evaluation of Speech Quality (PESQ) - A New Method for Speech Quality Assessment of Telephone Networks and Codecs. *IEEE International Conference on Acoustics, Speech, and Signal Processing*.
- [12] C. H. Taal, R. C. Hendriks, R. Heusdens, J. Jensen. (2011, feb.). An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech. *IEEE Transactions on Audio, Speech, and Language Processing*. vol. 19, págs. 2125-2136.
- [13] J. Heymann, L. Drude, A. Chinaev, R. Haeb-Umbach. (2015, dic.). BLSTM supported GEV beamformer front-end for the 3RD CHiME challenge. *IEEE Workshop on Automatic Speech Recognition and Understanding*.
- [14] M. Taseska, E. Habets. (2017, nov.). Nonstationary Noise PSD Matrix Estimation for Multichannel Blind Speech Extraction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. vol. 25, págs. 2223-2236.
- [15] D. P. Kingma, J. Ba. (2014, dic.). Adam: A Method for Stochastic Optimization. *3rd International Conference for Learning Representations*.



**Ángel Manuel Gómez García** es Licenciado y Doctor en Ingeniería Informática por la Universidad de Granada, España, en 2001 y 2006, respectivamente. En 2002 se unió al Departamento de Teoría de la Señal, Telemática, y Comunicaciones de la Universidad de Granada, donde es miembro del grupo de investigación en Procesado de Señal, Transmisión Multimedia y Tecnologías del Habla/Audio (SigMAT). Actualmente es un profesor titular en la Universidad de Granada. Sus intereses investigadores incluyen procesamiento del habla humana y aprendizaje automático (machine learning).



**Antonio Miguel Peinado Herreros** es Licenciado y Doctor en Ciencias Físicas (especialidad Electrónica) por la Universidad de Granada en 1987 y 1994, respectivamente. En 1988 trabajó como Ingeniero de Control de Calidad en la empresa pública Inisel. Desde entonces, ha desarrollado su labor en la Universidad de Granada (UGR), donde ha participado y liderado numerosos proyectos de investigación en el campo del procesado y transmisión de señales. En 1989, fue Consultant en el departamento de Speech Research de AT&T Bell Labs, Murray Hill, NJ, USA, y en 2018 fue Visiting Scholar en el Language Technologies Institute of CMU, Pittsburgh, PA, USA. Dentro de la UGR, ha sido Profesor Titular de Universidad desde 1996 hasta 2010, y desde entonces Catedrático de Universidad en el Dpto. de Teoría de la Señal, Telemática y Comunicaciones, donde actualmente dirige el grupo de investigación en Signal Processing, Multimedia Transmission and Speech/Audio Technologies. Es autor de numerosos artículos en revistas y congresos, así como autor del libro *Speech Recognition Over Digital Channels* (New York, NY, USA: Wiley, 2006).

Actualmente, sus intereses investigadores se centran en diversas tecnologías de voz (antispoofting para autenticación, realce, reconocimiento y transmisión), procesamiento de imágenes y procesado de señales proteómicas. Ha sido revisor de diversas revistas especializadas y de numerosos congresos, evaluador de proyectos y becas, y miembro del Comité Técnico de Programa de varias conferencias internacionales.



**David Pareja Martos** obtuvo el Grado en Ingeniería de Tecnologías de Telecomunicación en la Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada, en 2016. Ha obtenido en 2019 el Máster Universitario en Ingeniería de Telecomunicación, también en la Universidad de Granada.

Actualmente se halla trabajando en Telmi Telecom S.L., operadora de telecomunicaciones con sede en la provincia de Granada.

# Detección de patologías degenerativas mediante voz

Autor: Miriam González Atienza, e-mail: myriamgonzalez@correo.ugr.es

Tutores: José A. González López, e-mail: joseangl@ugr.es

Antonio M. Peinado Herreros, e-mail: amp@ugr.es

Titulación: Ingeniería de Telecomunicación

Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—En este proyecto, se abordará el problema de la detección automática de demencia y de la enfermedad de Alzheimer, a partir de los parámetros acústicos y lingüísticos extraídos de grabaciones de voz de los pacientes y de sus correspondientes transcripciones. Para obtener una caracterización lingüística adicional, se implementa un conjunto de algoritmos de extracción de parámetros basados en redes neuronales (*word embeddings*). El sistema de detección incluye una etapa de procesado de la señal y otra de clasificación basada en técnicas de aprendizaje computacional, como máquinas de vectores de soporte o redes neuronales profundas. Las métricas de clasificación más elevadas se han obtenido con la base de datos DementiaBank, con selección de parámetros acústicos y con el método de clasificación *Adaptive Boosting*, alcanzando un valor de precisión del 97.4%. Por lo tanto, se ha conseguido desarrollar una herramienta automática de detección de demencia a partir de parámetros acústicos y lingüísticos.

**Palabras clave**—Análisis acústico de la voz, aprendizaje computacional, demencia, diagnóstico de enfermedades, enfermedad de Alzheimer, redes neuronales profundas

## I. INTRODUCCIÓN

La demencia es un desorden neurodegenerativo causado principalmente por la EA (enfermedad de Alzheimer). Se trata de un trastorno progresivo que provoca la degeneración de las células del cerebro.

La parte del córtex cerebral que procesa las habilidades del lenguaje es una de las primeras en ser afectadas. Por lo tanto, uno de los síntomas cognitivos de la enfermedad de Alzheimer es la pérdida de la capacidad para expresarse y comprender el lenguaje. Aloisius Alzheimer describió que los pacientes diagnosticados con EA presentan parafasia, pausas en el habla y deficiencias en la comprensión lectora y escrita, además de pérdida de memoria. Normalmente, las alteraciones del lenguaje verbal se manifiestan de forma simultánea con las del lenguaje escrito.

Actualmente, no existe una herramienta eficaz que proporcione un diagnóstico fiable de la demencia. El paciente debe pasar por una serie de tests cognitivos llevados a cabo por un neurólogo, proceso que puede provocar ansiedad y estrés. Especialmente en el caso de un estado de detección temprana, estos métodos son invasivos, relativamente costosos y requieren una cantidad significativa de esfuerzo y tiempo. Encontrar un método de diagnóstico no invasivo que proporcione comodidad al paciente es, por lo tanto, de gran interés.

El objetivo general de este proyecto es el desarrollo de una herramienta automática de detección de la demencia mediante grabaciones de voz realizadas a sujetos de esta enfermedad y controles sanos.

La estructura del sistema de detección es la siguiente: una primera fase de preprocesamiento de las señales (locuciones y transcripciones), una fase de extracción de parámetros acústicos y lingüísticos con la correspondiente selección de aquellos más relevantes, una etapa de modelización con distintos tipos de clasificadores, y una etapa final de evaluación del sistema propuesto. La etapa de clasificación se lleva a cabo en dos bases de datos, en inglés (DementiaBank) [1] y en español, las cuales contienen grabaciones de pacientes con EA y sujetos de control sanos. Se ha escogido trabajar con la base de datos DementiaBank por ser de las más utilizadas en la literatura, [2], [3], [4], [5] y con el objetivo de tener una referencia de los resultados obtenidos.

## II. CONTRIBUCIONES

Respecto a los trabajos presentes en la literatura, este proyecto aporta una nueva combinación de parámetros acústicos y lingüísticos descritos en los siguientes capítulos, así como la extracción de *word embeddings* mediante redes neuronales y su posterior utilización como parámetros lingüísticos. Además, se ha conseguido aumentar la precisión en la clasificación respecto a la obtenida para la base de datos DementiaBank, en [2], [3], [6]. Como novedad, también se ha utilizado una base de datos en español con un menor número de sujetos que DementiaBank, para comparar los resultados obtenidos con ambas. Existen muy pocos trabajos que hayan trabajado con sujetos cuya lengua nativa es el español, ya que la mayoría de trabajos son en inglés o alemán.

## III. BASES DE DATOS

Se han utilizado dos bases de datos: una en español y otra en inglés.

Para elaborar la base de datos en español, el estudio fue realizado de forma prospectiva, entre marzo y abril de 2018, en el centro Gerontológico Ancha de Capuchinos y en el centro La Chana. Los participantes aceptaron realizar el estudio de forma voluntaria bajo consentimiento informado.

En este estudio participaron 20 sujetos, 10 de ellos con deterioro cognitivo que formaban el grupo experimental y 10 sanos que componían el grupo de control.

La tabla I muestra los detalles demográficos de los sujetos de la base de datos en español.

La segunda base de datos que se ha utilizado es DementiaBank [1], creada por la Universidad de Medicina de Pittsburgh. Esta base de datos contiene grabaciones de voz

Tabla I  
DETALLES DE LA BASE DE DATOS EN ESPAÑOL

Variables	Grupo EA	Grupo control
Mujeres	8	9
Hombres	2	1
Rango edad	[80,97] años	[78,89] años
Duración grabaciones	17.45 minutos	20.15 minutos

mientras los participantes del estudio describían la imagen del robo de las galletas del test de Boston [7].

La tabla II recoge las características de la base de datos DementiaBank.

Tabla II  
DETALLES DE LA BASE DE DATOS DEMENTIABANK

Variables	Grupo EA	Grupo control
Número sujetos	167	97
Rango edad	[49,90] años	[46,81] años
Duración grabaciones	160 minutos	92 minutos

#### IV. PRE-PROCESAMIENTO

Debido al ruido de fondo presente en las grabaciones de las bases de datos, es necesario aplicar una técnica de reducción de ruido, con el objetivo de evitar posibles errores en la etapa de clasificación.

En este caso, la técnica elegida es *Noise reduction using spectral gating*, utilizada en [2].

En la figura 1 se representa el resultado de aplicar esta técnica a una de las grabaciones de los sujetos de la base datos.

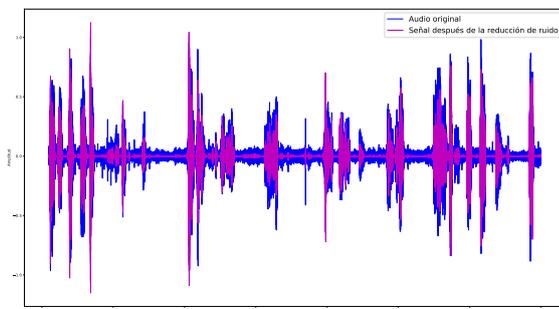


Fig. 1. Señal de audio tras aplicarle la reducción de ruido.

#### V. MÉTODO DE DETECCIÓN PROPUESTO

La figura 2 muestra los pasos seguidos en el sistema de detección de demencia.

En primer lugar, se aplica una etapa de preprocesamiento a la señal de entrada. Esta etapa depende del tipo de parámetros (acústicos o lingüísticos). Los parámetros acústicos se extraen a partir de la señal de voz, por lo que esta etapa consiste en un filtro de eliminación de ruido. Por otra parte, los parámetros lingüísticos se extraen a partir de las transcripciones de las grabaciones, por lo que esta etapa consiste en la selección de la parte en la que habla el sujeto (de control y del grupo EA), así como en la eliminación de caracteres que puedan dar lugar

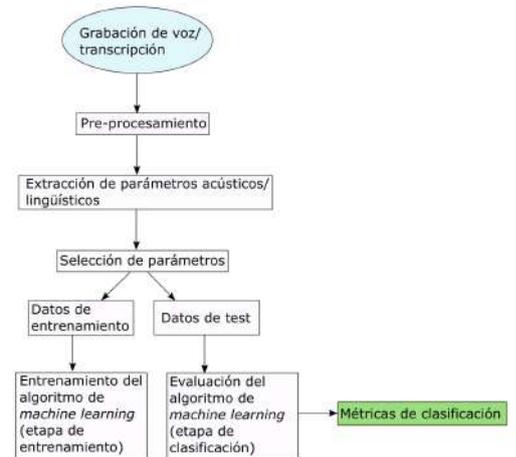


Fig. 2. Metodología seguida

a un error en la extracción de parámetros. Tras esta etapa, se construye la matriz de parámetros y se aplica una técnica de selección de aquellos más relevantes. Finalmente, se aplica el método de clasificación (ver figura 2).

Los parámetros acústicos están relacionados con la forma de hablar de un sujeto, mientras que los parámetros lingüísticos se relacionan con el contenido del discurso.

Para evaluar los algoritmos de detección de la demencia implementados, es necesario disponer de un amplio conjunto de datos que dividiremos en entrenamiento y test. Este conjunto de datos está formado por los parámetros acústicos y lingüísticos extraídos de cada sujeto. A continuación se describen estos parámetros, el criterio para su elección y su influencia en la etapa de clasificación.

##### A. Parámetros acústicos

Los parámetros acústicos extraídos se dividen en las siguientes categorías:

- **Parámetros prosódicos.** Los parámetros como la duración de las pausas o del habla se han extraído a partir de la salida del detector de voz (VAD).

En primer lugar, se aplica una ventana a la señal de 30 ms de duración con un 50% de solapamiento.

Para determinar si un segmento de la señal es de voz o silencio, se utiliza un criterio de energía. La hipótesis de partida es que una secuencia de voz de larga duración tiene un ratio de energía superior a un cierto umbral, establecido como el 60% de la energía de la señal.

Finalmente, se aplica el criterio de energía obteniendo una clasificación entre segmentos de voz y de silencio. La salida es binaria, es decir, toma el valor "1" si se ha detectado voz y el valor "0" si se ha detectado silencio. A partir de la salida del VAD se han determinado 4 parámetros acústicos: duración de los segmentos de voz, duración de los segmentos de silencio (pausas), proporción entre las pausas y la duración total del audio y proporción entre las pausas y el habla.

La figura 3 muestra el resultado de aplicar el VAD implementado sobre una de las señales de audio:

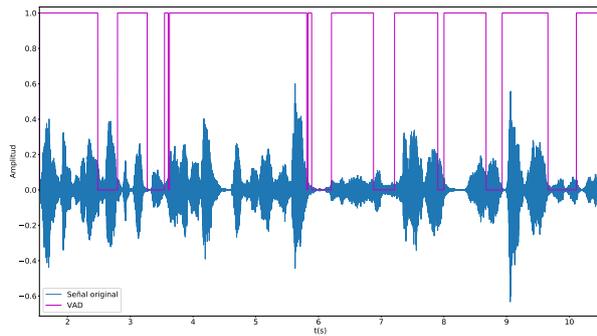


Fig. 3. Salida del VAD sobre una señal de audio.

La información extraída del VAD resulta especialmente relevante en el proceso de clasificación, ya que es capaz de proporcionar una clara distinción entre personas con EA, y pertenecientes al grupo de control. Las características que manifiesta el lenguaje de las personas con EA pueden variar drásticamente de un paciente a otro, de acuerdo con la progresión de la enfermedad.

Además de la longitud de las pausas y la longitud del habla, también se ha considerado relevante extraer parámetros como la proporción entre la longitud de las pausas y la longitud total de la locución, y la proporción entre la duración de las pausas y la duración del habla. Esto se debe a que la mayoría de personas con EA realizan un mayor número de pausas en su discurso que las personas del grupo de control.

Otro parámetro prosódico extraído ha sido el ritmo del habla en bpm (*beats per minute*). Para ello, se ha implementado el método PLP (*Predominant Local Pulse estimation*), consistente en analizar la envolvente de la señal en el dominio de la frecuencia para encontrar un tiempo (o ritmo) estable en cada trama.

- **MFCCs (Mel Frequency Cepstral Coefficients)**  
Se han extraído 42 coeficientes cepstrales. Los MFCCs se han calculado a nivel de frase, es decir, únicamente en los segmentos de voz, ya que calcularlos sobre los segmentos de pausas no aporta información útil. Además, se ha calculado su asimetría, curtosis, media y varianza.
- **Centroides espectrales.** Se han extraído 42 centroides cepstrales en total.
- **Jitter.** Se define como la variabilidad temporal de la frecuencia fundamental ( $F_0$ ).
- **Pitch.** Para la extracción del pitch, se ha utilizado el método de autocorrelación. En primer lugar, se define un intervalo de búsqueda del *pitch*. Este rango varía entre hombres y mujeres, siendo entre  $77\text{ Hz}$  y  $482\text{ Hz}$  para hombres y entre  $137\text{ Hz}$  y  $634\text{ Hz}$  para mujeres. Por lo tanto, en el algoritmo se ha establecido el rango  $[65, 500]\text{ Hz}$ . El siguiente paso es localizar los picos de la señal en ese rango y localizar el máximo de ellos. Este máximo se corresponde con el *pitch*.

## B. Parámetros lingüísticos

Como se ha mencionado previamente, las capacidades comunicativas de las personas con EA se ven afectadas en las primeras fases de la enfermedad. Por esta razón, resulta relevante el estudio de estos parámetros para la fase de detección.

Es importante destacar que los parámetros expuestos se han calculado en proporción con el número total de palabras, ya que la longitud del discurso varía según el sujeto, y este es un parámetro de gran relevancia. Para obtener este parámetro de normalización (número total de palabras de la transcripción de cada locución), se realiza un conteo del número de palabras totales, es decir, se obtiene la longitud del discurso del participante en número de palabras. Este parámetro servirá para obtener una proporción de los demás parámetros. Por ejemplo, un paciente perteneciente al grupo EA puede decir un elevado número de palabras, y, de estas, solamente un 5% serán únicas. Sin embargo, un sujeto perteneciente al grupo de control puede tener un discurso de menor longitud y, sin embargo, la proporción de palabras únicas es superior. Por lo tanto, para evitar errores en la clasificación, se han extraído los parámetros lingüísticos en forma de proporción (excepto para el número de palabras).

Para extraer los parámetros lingüísticos se ha utilizado la herramienta de Python *NLTK* (*Natural Language Toolkit*). Constituye una herramienta útil para el procesamiento de lenguaje, permitiendo categorizar textos, analizar su estructura sintáctica, y otras utilidades.

Los parámetros lingüísticos extraídos han sido los siguientes:

- **Proporción entre el número de palabras únicas y el número de palabras totales:**

La figura 4 muestra el número medio de palabras totales y de palabras únicas registradas en las transcripciones de una muestra de sujetos de ambos grupos, de la cual se puede concluir que, como era previsible, los sujetos del grupo de control realizan un discurso más extenso que los sujetos del grupo experimental.

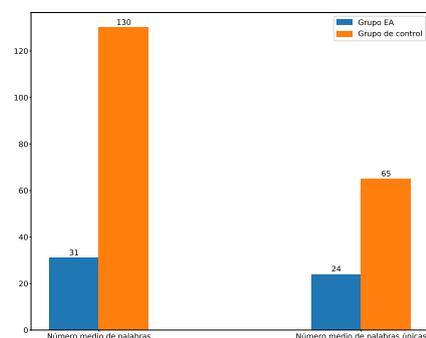


Fig. 4. Número medio de palabras totales y de palabras únicas, grupo de control y grupo EA.

Los sujetos pertenecientes al grupo experimental, por lo tanto, dicen un menor número de palabras únicas en su discurso, como se puede ver en la gráfica 4, ya que una

de las características de la enfermedad es la repetitividad en el lenguaje.

- **Etiquetas POS (*Part of speech*).**

El proceso mediante el cual se clasifican las palabras en sus correspondientes clases gramaticales se conoce como *part-of-speech tagging*. Para la extracción de estas clases gramaticales (etiquetas POS), se ha utilizado la librería de Python *NLTK*, previamente mencionada.

A partir de las etiquetas POS, se puede calcular el número total de verbos, determinantes, sustantivos, pronombres,.. presentes en cada transcripción. Se ha calculado la proporción total de nombres, verbos, determinantes y pronombres respecto del número total de palabras. Es decir, para cada transcripción, se divide el número total de verbos (igualmente con las demás categorías) entre el número total de palabras.

### C. Word embeddings

Se trata de un método de extracción de parámetros alternativo al uso de parámetros acústicos y lingüísticos, por utilizar redes neuronales en el contexto de los métodos de clasificación.

*Word embedding* hace referencia a conjunto de técnicas de modelado de lenguaje en las que las palabras de un determinado vocabulario son representadas mediante vectores de números reales y de longitud fija.

El modelo utilizado para dicha representación ha sido *Word2vec*, debido a su eficiencia computacional. Este modelo se encuentra disponible de dos formas: *CBOW* o *Skip-Gram*. En el modelo *Skip-Gram*, se parte de un conjunto de frases (*corpus*) y se analizan las palabras que las forman para tratar de predecir las palabras vecinas. En el modelo *CBOW*, se combinan las distintas representaciones del contexto de cada palabra para predecir la palabra central. La principal diferencia entre ambos métodos se encuentra en el objetivo final de la predicción. Mientras que en *CBOW* se utiliza el contexto para predecir la palabra central, en el modelo *Skip-Gram* se utilizan las representaciones de las palabras para predecir el contexto. Este último suele proporcionar mejores resultados de clasificación que *CBOW*.

Ambas arquitecturas están diseñadas para aprender la representación de cada palabra haciendo uso de redes neuronales. La figura 5 muestra un esquema del funcionamiento de cada modelo donde se aprecia la diferencia entre la salida proporcionada por ambos.

A partir de la capa oculta del modelo se obtienen los *word vectors*. Estos son sus pesos. Es importante definir el número de neuronas de esta capa. En este caso, se han entrenado vectores de palabras con 300 características. Por lo tanto, la matriz de representación tendrá dimensión (longitud vocabulario, 300).

El modelo *Skip-Gram* proporciona mejores resultados para palabras poco frecuentes en el vocabulario, al contrario que *CBOW*. En este último, el modelo trata de predecir la palabra dado su contexto, es decir, encontrar la palabra con probabilidad máxima de aparición en el contexto. El modelo asigna una baja probabilidad a las palabras poco frecuentes, ya que no aparecen muy a menudo en un contexto dado. Por otra parte, *Skip-Gram* trata de predecir el contexto dada una palabra.

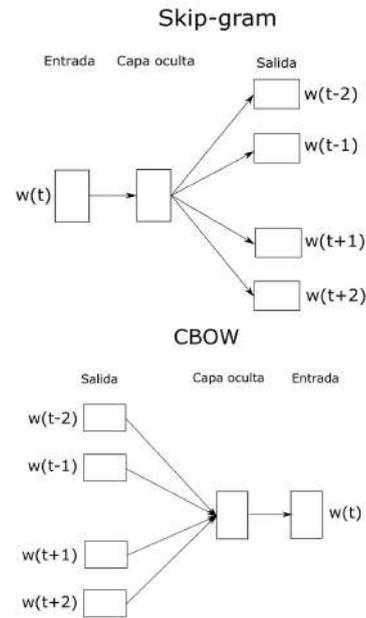


Fig. 5. Esquema de Skip-Gram y CBOW

En este caso, una palabra frecuente y otra no frecuente son tratadas del mismo modo: como palabra y como contexto. Es decir, el modelo aprende incluso las palabras que no suelen aparecer.

### D. Selección de parámetros

Debido al alto número de parámetros extraídos, es necesario seleccionar cuáles de estos resultan más relevantes para la etapa de clasificación y proporcionan unos mejores resultados. De esta forma, se acelera el tiempo dedicado a la etapa de clasificación y se consigue evitar el fenómeno de sobreentrenamiento.

Para visualizar la importancia de cada parámetro, se ha implementado la técnica *ExtraTreesClassifier* [8]. Consiste en un modelo de clasificación basado en árboles de decisión. La repercusión de los parámetros en la etapa de clasificación se puede medir como su relevancia en la decisión tomada por cada árbol. Se representa mediante un número comprendido entre 0 y 1 para cada parámetro, donde 0 significa que no se utiliza para la predicción y 1 significa que es capaz de predecir correctamente las muestras de test. La importancia de todos los parámetros siempre suma 1.

## VI. RESULTADOS GENERALES

Para llevar a cabo la etapa de detección, se han implementado los métodos de clasificación más relevantes en la literatura: Máquinas de vectores de soporte (SVM), Bosques aleatorios (*Random forests*), *Adaptive Boosting*, y redes neuronales profundas (DNNs). La razón por la cual se han escogido estos algoritmos de clasificación es por su amplio uso en la literatura, así como por los elevados resultados de precisión que ofrecen.

Además, para la evaluación de estos algoritmos, se ha utilizado la técnica *cross-validation*, en la que se construyen  $N$  diferentes modelos y las predicciones se calculan sobre

todos los datos, permitiendo utilizar todas las muestras como entrenamiento y test. Por lo tanto, los resultados obtenidos son más concluyentes debido a que se toma la media de la precisión de todos los modelos empleados.

Las métricas de clasificación calculadas han sido las siguientes:

- **Sensibilidad (*recall*):** Indica la capacidad del clasificador para dar como positivos los casos realmente positivos. Es decir, indica la proporción de enfermos correctamente identificados.
- **Especificidad (*specificity*):** Indica la capacidad del clasificador para dar como casos negativos los casos realmente negativos. Es decir, indica la proporción de sanos correctamente identificados.
- **Precisión (*accuracy*):** Se calcula como la proporción entre el número de predicciones correctas y el número total de predicciones.

Un sistema con un elevado porcentaje de sensibilidad y una baja precisión devuelve una gran cantidad de resultados, pero la mayoría de las etiquetas predichas son incorrectas. Por otra parte, un sistema con una elevada precisión y baja sensibilidad devuelve pocos resultados, pero la mayoría de las etiquetas predichas son correctas.

#### A. Combinación de parámetros

En la figura 6 se pueden visualizar los resultados de clasificación combinando parámetros acústicos y lingüísticos, con la base de datos DementiaBank.

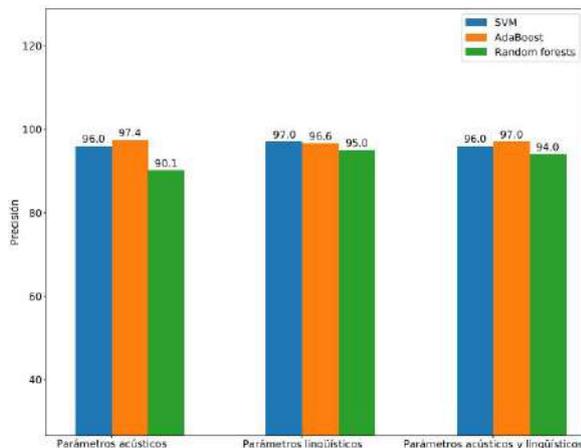


Fig. 6. Resultados de precisión con selección de parámetros acústicos y lingüísticos para DementiaBank

De la figura 6 se puede concluir que la combinación de parámetros acústicos y lingüísticos, en el caso del clasificador *Adaptive boosting*, supone una mejora respecto a la utilización únicamente de parámetros lingüísticos en términos de precisión.

Sin embargo, se puede observar que, por ejemplo, utilizando SVM y Bosques aleatorios, una combinación de parámetros acústicos y lingüísticos no supone una mejora significativa en los resultados de clasificación.

El máximo valor de precisión alcanzado con la base de datos DementiaBank, ha sido con *Adaptive boosting*, utilizando únicamente selección de parámetros acústicos. Se ha

alcanzado un valor de precisión del 97.4%, un 93.32% de especificidad y un 100% de sensibilidad.

#### B. Parámetros lingüísticos junto con word embeddings

Para evaluar los resultados de clasificación de la red neuronal (DNN), se han utilizado los parámetros lingüísticos extraídos junto con los *word embeddings*. La tabla III muestra los resultados de precisión/especificidad/sensibilidad para las distintas combinaciones.

Tabla III  
RESULTADOS DE CLASIFICACIÓN CON *word embeddings*, PARA LA BASE DE DATOS DEMENTIABANK

Lingüísticos con Skip-gram	Lingüísticos con CBOW	Lingüísticos
99.93/94.75/99.4	99.77/93.21/97.6	94.13/84.30/96.14

De la tabla III se puede concluir que la combinación de parámetros lingüísticos con el modelo de *word embeddings* Skip-Gram para extraer los parámetros característicos de cada grupo, y una posterior clasificación mediante una red neuronal de 300 neuronas y *backpropagation*, ofrece un mayor valor de precisión, sensibilidad y especificidad que utilizando únicamente parámetros acústicos y lingüísticos o parámetros lingüísticos sin *word embeddings* o con el modelo CBOW.

## VII. DISCUSIÓN

Los mejores resultados de clasificación obtenidos para la base de datos DementiaBank han sido con el clasificador *Adaptive boosting*, y con selección de parámetros acústicos, con una precisión del 97.4%, especificidad del 93.32 % y sensibilidad del 100 %. Los parámetros acústicos seleccionados han sido: Media de los 42 primeros MFCCs, ritmo del habla, duración del habla, duración de las pausas, varianza del pitch, varianza de las derivadas de los 42 primeros MFCCs.

Por otra parte, los mejores resultados de clasificación obtenidos con la base de datos en español han sido con el clasificador *Adaptive boosting* y con selección de parámetros acústicos y lingüísticos, con una precisión del 97.2%, especificidad del 94.44 % y sensibilidad del 100 %. Los parámetros acústicos y lingüísticos seleccionados han sido: tasa de fluidez fonológica, tasa de fluidez semántica, media de los 42 primeros MFCCs, duración del habla, ritmo del habla, varianza de los 42 primeros MFCCs, y proporción entre las pausas y la longitud de la locución.

La tabla IV recoge una comparación entre los resultados obtenidos por los estudios más relevantes de la literatura y los obtenidos en este proyecto.

Tabla IV  
COMPARACIÓN DE LOS RESULTADOS OBTENIDOS CON LOS DE LA LITERATURA, PARA LA BASE DE DATOS DEMENTIABANK.

Estudio	Precisión alcanzada	Precisión alcanzada en este proyecto
[2]	91.96 % ( <i>Random forests</i> ) 85.83 % (ADA)	90.10 % ( <i>Random forests</i> ) 97.40 % (SVM)
[3]	85.9 % (SVM)	96 % (SVM)
[9]	92.5 % (RNN)	99.77 % (DNN)

### VIII. CONCLUSIONES

En este trabajo se proponen dos métodos automáticos de clasificación que pueden ser utilizados como método de diagnóstico alternativo de demencia utilizando únicamente la voz del paciente.

- Técnicas de aprendizaje computacional (SVM, *Adaptive boosting* y *Random Forests*), cuyos datos de entrenamiento son los parámetros acústicos y lingüísticos extraídos de las grabaciones de los participantes. Con estos métodos de clasificación, se ha conseguido obtener unos resultados de precisión que, en el caso de SVM y *Random Forests*, son superiores a los alcanzados en la literatura. El máximo de precisión obtenido ha sido con el clasificador *Adaptive boosting*, haciendo uso únicamente de parámetros acústicos.
- Redes neuronales profundas, cuyo conjunto de entrenamiento consiste, o bien en los parámetros extraídos de los sujetos, o bien en la combinación de parámetros lingüísticos junto con los *word embeddings* extraídos a partir de las transcripciones de cada grabación. Con este método se ha conseguido alcanzar un valor de precisión del 99.93% combinando parámetros lingüísticos junto con *word embeddings* extraídos mediante el modelo Skip-Gram.

Por lo tanto, podemos concluir que se ha conseguido desarrollar una herramienta automática para detectar la demencia causada por la EA a partir de parámetros acústicos y lingüísticos extraídos a partir de grabaciones de voz de personas que padecen esta enfermedad. Igualmente, se ha obtenido una caracterización lingüística del discurso de los participantes mediante *word embeddings*, parámetros que han demostrado proporcionar unos elevados resultados de precisión en la clasificación. Además, se ha logrado superar, en la mayoría de los casos, los resultados de clasificación presentes en la literatura.

### IX. TRABAJO FUTURO

En el futuro, sería interesante implementar el método de clasificación descrito en [4]. Éste se basa en la creación de un modelo de lenguaje basado en red neuronal (NNLMs) con celdas de tipo LSTM (Long Short Term-Memory), diseñado para clasificar sujetos sanos y con EA.

Asimismo, se planea investigar un mayor conjunto de parámetros acústicos, como por ejemplo: i-vector [10], shimmer o HNR, así como un mayor conjunto de parámetros lingüísticos, por ejemplo: riqueza léxica utilizando la medida *Brunet's W index* y *Honore's R Statistics* o parámetros de perplejidad [10], [6].

Finalmente, la aceleración computacional pasará también por la paralelización en memoria distribuida de los algoritmos y de las librerías de base, para lo que se considerarán lenguajes de más bajo nivel.

### AGRADECIMIENTOS

En primer lugar, quiero agradecer este trabajo a mis padres, por su cariño y apoyo incondicional durante toda la carrera. Sin ellos, no habría llegado a donde estoy ahora. Por todo lo que me han enseñado y lo que me han dado.

A mis tutores, Antonio y José Andrés, por orientarme en este proyecto, por su apoyo y disponibilidad siempre que lo he necesitado, y por sus excelentes labores docentes y de tutelaje.

### REFERENCIAS

- [1] [Online]. Available: <https://talkbank.org/DementiaBank/>. [Accessed: 10-Dec-2015]DementiaBank.[Online].Available:<https://talkbank.org/DementiaBank/>
- [2] S. Al-Hameed, M. Benaissa, and H. Christensen, "Simple and robust audio-based detection of biomarkers for alzheimer's disease."
- [3] —, "Detecting and predicting alzheimer's disease severity in longitudinal acoustic data," 2017.
- [4] J. Fritsch, S. Wankerl, and E. Nöth, "Automatic diagnosis of alzheimer's disease using neural network language models," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 5841–5845.
- [5] B. Mirheidari, D. Blackburn, M. Reuber, T. Walker, and H. Christensen, "Diagnosing people with dementia using automatic conversation analysis."
- [6] J. Weiner, M. Engelbart, and T. Schultz, "Manual and automatic transcriptions in dementia detection from speech," in *Proc. Interspeech 2017*, 2017, pp. 3117–3121. [Online]. Available: <http://dx.doi.org/10.21437/Interspeech.2017-112>
- [7] W. J. Mack, D. M. Freed, B. W. Williams, and V. W. Henderson, "Boston naming test: Shortened versions for use in alzheimer's disease," vol. 47, pp. P154–P158, 1992.
- [8] [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html>
- [9] B. Mirheidari, D. Blackburn, T. Walker, A. Venneri, M. Reuber, and H. Christensen, "Detecting signs of dementia using word vector representations."
- [10] J. Weiner and T. Schultz, "Selecting features for automatic screening for dementia based on speech," pp. 747–756, 2018.

# Técnicas de Anti-Spoofing Robustas al Ruido para Verificación Automática de Locutores

Autor: Alejandro Gómez Alanís, e-mail: agomezalanis@ugr.es  
Tutor: Antonio Miguel Peinado Herreros, e-mail: amp@ugr.es  
Tutor: José Andrés González López, e-mail: j.gonzalez@uma.es  
Titulación: Máster en Ingeniería de Telecomunicación  
Departamento de Teoría de la Señal, Telemática y Comunicaciones  
Universidad de Granada

**Resumen**—Los sistemas de verificación automática de locutores (ASV) han incrementado su popularidad en los últimos años debido a que empiezan a ser introducidos de forma habitual en sistemas comerciales de autenticación biométrica basada en la voz del usuario. Estos sistemas, no obstante, presentan vulnerabilidades en las que un impostor podría hacerse pasar por un usuario legítimo: estos son los denominados ataques de suplantación o *spoofing*. Actualmente, existen herramientas que permiten generar voz suplantada con relativos poco conocimientos técnicos, tal y como ocurre con los ataques de síntesis y conversión de voz. A pesar de los esfuerzos dedicados hasta el momento, las técnicas de detección desarrolladas para este tipo de ataques disminuyen su rendimiento muy considerablemente en escenarios con distorsiones ruidosas y de reverberación. Con objeto de robustecer los sistemas ASV frente ataques de suplantación, en este trabajo se propone una nueva técnica de detección basada en redes neuronales profundas, que permite clasificar la voz de entrada al sistema en genuina ó suplantada. Además, se propone una técnica basada en máscaras de ruido para robustecer al sistema frente a distorsiones ruidosas. La evaluación del sistema se realiza sobre las versiones original y ruidosa de la base de datos ASVspoof2015. Los resultados experimentales demuestran la superioridad del sistema propuesto sobre el resto de sistemas de *anti-spoofing* de la literatura tanto en escenarios sin ruido como en escenarios con distorsiones ruidosas y reverberantes.

**Palabras clave**—Verificación automática de locutores, detección de suplantación o *spoofing*, robustez al ruido, redes neuronales profundas, máscaras de ruido.

## I. INTRODUCCIÓN

EN los últimos años, la popularidad de los sistemas de verificación de locutores (ASV, del inglés *Automatic Speaker Verification*) se ha incrementado notablemente al empezar estos a ser introducidos de forma habitual en sistemas comerciales [1]. ASV es una técnica de identificación biométrica en la que se pretende detectar a través de la voz si la identidad reclamada por un locutor es cierta [2]. Sin embargo, la mayoría de sistemas ASV actuales no integran medidas para detectar impostores que intentan acceder al sistema de forma fraudulenta, y existen evidencias de que estos sistemas son susceptibles de ataques maliciosos que pueden vulnerar su seguridad [3].

En un sistema ASV existen ocho puntos de vulnerabilidad, tal y como se muestra en la figura 1. Estas vulnerabilidades se pueden clasificar en dos tipos de ataques: (1) ataques directos o de suplantación, que son aquellos que se pueden aplicar tanto al nivel del micrófono como a nivel de transmisión de la

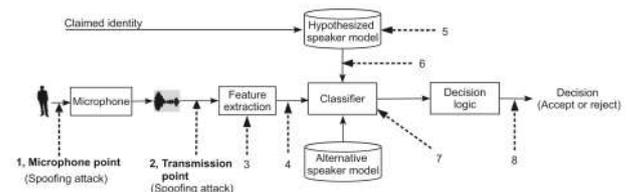
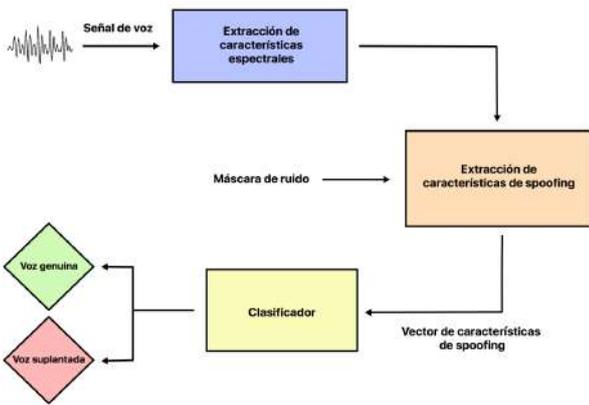


Fig. 1. Sistema de verificación automática (ASV) con 8 posibles puntos de ataque. Los ataques de los puntos 1-2 se consideran ataques directos o de *spoofing* mientras que los ataques de los puntos 3-8 son ataques indirectos [4].

señal, y (2) ataques indirectos, que son aquellos que requieren acceso al nivel del sistema ASV, y no suelen estar accesibles a los usuarios.

En ASV, existen cuatro tipos de ataques de suplantación [4]: (i) interpretación (imitación de la voz de un usuario legítimo), (ii) repetición (grabación y reproducción de la voz de un usuario legítimo), (iii) síntesis de voz (generación de voz artificial que se parece a la de un usuario legítimo), y (iv) conversión de voz (manipulación de voz para parecerse a la de un usuario legítimo). Esto ha conllevado al desarrollo de sistemas de *anti-spoofing* para discriminar entre voz genuina y voz suplantada, de manera que solo si el sistema de *anti-spoofing* detecta voz genuina, se procede a realizar la verificación del locutor. Este trabajo se centra en el desarrollo de contramedidas para los ataques de síntesis y conversión de voz, los cuales se pueden generar fácilmente con herramientas con relativo poco conocimiento técnico. En los últimos años, las técnicas de síntesis y conversión de voz han permitido obtener voz cuasi-realista, lo que supone un reto desde el punto de vista de la suplantación de voz [3].

El rendimiento de los sistemas de *anti-spoofing* desarrollados para detectar ataques en condiciones de voz limpia se reduce significativamente en entornos ruidosos [5], de forma que el deterioro incrementa a medida que la relación señal-ruido (SNR) entre la voz y el ruido acústico disminuye. En este trabajo se propone un sistema de *anti-spoofing* basado en redes neuronales profundas, que permite clasificar la voz de entrada al sistema ASV como voz genuina ó suplantada. Además, se propone una nueva metodología para contrarrestar las distorsiones ruidosas basada en máscaras de ruido [6], que informan al sistema cuáles son los bins tiempo-frecuencia que están más contaminados por ruido. Por otro lado, se realiza un estudio del rendimiento de diferentes características

Fig. 2. Diagrama de bloques del sistema de *anti-spoofing* propuesto.

espectrales y clasificadores finales para elegir la combinación que ofrece un mayor rendimiento en la detección de ataques de suplantación.

## II. SISTEMA DE ANTI-SPOOFING PROPUESTO

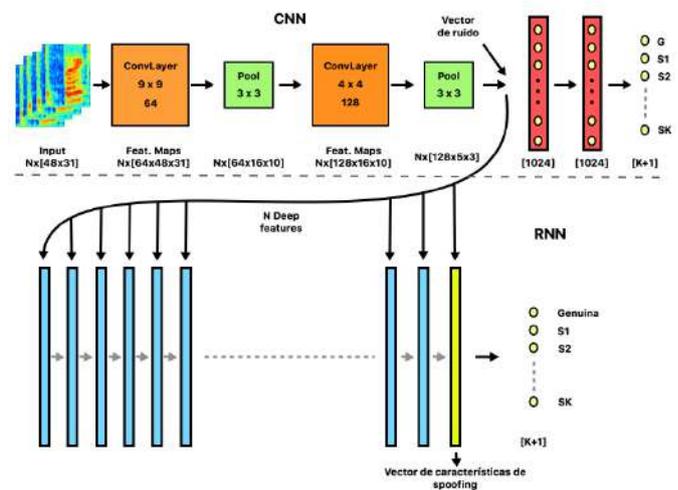
En la figura 2 se muestra un diagrama de bloques con las partes fundamentales que componen el sistema de *anti-spoofing* propuesto. En primer lugar, se extraen características espectrales de la señal de voz que alimentan al sistema. A continuación, se extraen características de suplantación de la señal de voz, teniendo en cuenta las posibles distorsiones presentes en la señal. Una vez que se obtiene el vector de características de *spoofing* de la elocución, el clasificador del sistema se encarga de discernir entre voz genuina ó suplantada.

### A. Extracción de características espectrales de la voz

En este trabajo se evalúan tres tipos de parametrizaciones en tiempo-frecuencia de la señal de voz: análisis FBANK, CQCC y MFCC. Las características FBANK se obtienen aplicando la Transformada de Fourier de Tiempo Corto (STFT) y realizando un filtrado al logaritmo del espectro en magnitud mediante un banco de  $M$  filtros en escala Mel. Sin embargo, estas características suelen estar altamente correlacionadas. Una forma de decorrelar estas características es aplicando la Transformada Discreta del Coseno (DCT) para obtener las clásicas características MFCC (*Mel Frequency Cepstral Coefficients*).

En los últimos años se han vuelto muy populares las características CQCC (*Constant Q Cepstral Coefficients*) para la detección de suplantación [9]. Se obtienen aplicando la Transformada de Q Constante (CQT), donde el factor  $Q$  es una medida de la selectividad de cada filtro, y se define como la razón entre la frecuencia central y el ancho de banda del filtro.

A diferencia de la STFT, cuyo factor  $Q$  incrementa al aumentar la frecuencia debido a que el ancho de banda es el mismo para todos los filtros, el ancho de banda de los filtros empleado en la CQT no es constante. Esto implica que la CQT presenta una mayor resolución en frecuencia en las frecuencias bajas, y una mayor resolución temporal en las frecuencias altas.

Fig. 3. Arquitectura de la red neuronal del sistema de *anti-spoofing* que extrae un vector de características por cada elocución ( $N$  representa el número de ventanas de contexto que componen una señal de voz).

### B. Extracción de características de suplantación

La arquitectura que se propone para extraer características de *spoofing* se muestra en la figura 3. Una ventana de contexto compuesta por  $W$  tramas temporales consecutivas, centrada en la trama que se está procesando en esa iteración, se usa para obtener las características espectrales que alimentan al sistema. Después, se emplea una red neuronal convolucional (CNN) para procesar las características espectrales y obtener características de mayor nivel de abstracción (*deep features*). En concreto, la CNN obtiene un vector de 1920 componentes por cada ventana temporal (*deep features*). Los vectores extraídos por la CNN son procesados y agregados por una red neuronal recurrente (RNN), que representa la elocución completa mediante un único vector de características.

En esta arquitectura, la CNN actúa como un extractor de características de suplantación a nivel de trama temporal, obteniendo un vector de características (*deep features*) por cada ventana de contexto de  $W$  tramas temporales consecutivas. Para ello, la CNN se implementa como un clasificador cuyo objetivo consiste en determinar si cada trama temporal es voz genuina o si pertenece a uno de los  $K$  ataques de suplantación ( $S_1, S_2, \dots, S_K$ ) con los que se entrena el sistema. La CNN está compuesta por 2 capas convolucionales de 64 filtros de  $9 \times 9$  y de 128 filtros de  $4 \times 4$ . Además, a la salida de cada capa convolucional se emplea un filtro de *max-pooling* de  $3 \times 3$  para reducir la dimensión de cada salida, y estas capas convolucionales está conectadas con dos capas de 1024 neuronas sigmoideas totalmente conectadas.

Para prevenir el problema del sobreajuste durante el entrenamiento, se aplica una técnica de *annealed dropout* [10] en las capas totalmente conectadas que elimina un porcentaje de diferentes conexiones en cada iteración del entrenamiento. En esta técnica, la probabilidad de *dropout* de los nodos de la red se disminuye conforme progresa el entrenamiento. En este trabajo se utiliza una razón de *dropout* que parte de una razón inicial  $prob[0]$  hasta cero en  $N$  pasos con una razón constante. La probabilidad de *dropout*  $prob[t]$  en la época  $t$  es:

$$prob[t] = \max\left(0, 1 - \frac{t}{N}\right)prob[0]. \quad (1)$$

Tal y como se muestra en la figura 3, las características de *spoofing* obtenidas por la CNN a nivel de trama temporal, se procesan por una RNN para obtener el vector de características del conjunto de la señal de voz. La principal ventaja del uso de una RNN, basada en celdas GRU (*Gated Recurrent Units*) [11], es su capacidad para aprender dependencias temporales largas entre los diferentes vectores profundos de características (*deep features*) de las ventanas de contexto. El estado de la celda de la RNN en el último paso temporal representa el vector de características de *spoofing* de la elocución.

### C. Robustez al ruido

Uno de los principales objetivos de este trabajo es el desarrollo de un sistema de *anti-spoofing* para ASV que sea robusto a condiciones ruidosas y reverberantes. Para ello, se proponen dos técnicas complementarias: (i) un entrenamiento multi-condición (en diversos entornos ruidosos), y (ii) una técnica para contrarrestar el ruido basada en máscaras de ruido. En esta sección se denotan las características espectrales ruidosas como  $y$ , las características limpias desconocidas como  $x$ , y el ruido como  $n$ .

El entrenamiento de una red neuronal profunda con datos multi-condición permite que la red aprenda características de más alto nivel que son más invariantes a los efectos del ruido con respecto a la precisión de clasificación. En términos de ingeniería de características, las primeras capas de las redes profundas buscan implícitamente características discriminativas que son invariantes a lo largo de todas las condiciones acústicas presentes en los datos de entrenamiento.

La adaptación de modelos es una de las principales técnicas para conseguir robustez frente al ruido, ya que es capaz de superar el desajuste que existe entre las condiciones de entrenamiento y evaluación. Existe un modelo de distorsión (ecuación 2) para características espectrales logarítmicas [12], que puede usarse para implementar mecanismos de compensación, pero con el inconveniente de que al ser una expresión no lineal, resulta problemática para técnicas de compensación estadística clásicas.

$$y_t \approx x_t + \log(1 + e^{n_t - x_t}) \quad (2)$$

Uno de los principales retos para conseguir robustez frente al ruido en los sistemas de reconocimiento del habla es tratar esta relación no lineal. Sin embargo, en la arquitectura propuesta, las capas de procesamiento no lineal de la red neuronal profunda pueden tener la capacidad de aprender directamente esta relación a partir de los datos de entrenamiento. Es por ello que en este trabajo se propone una técnica para la adaptación de los modelos de suplantación al entorno acústico a partir de cierta información sobre las características del ruido acústico presente en este entorno: esto es lo que se conoce como entrenamiento *noise-aware* en la literatura [8].

Para implementar el entrenamiento *noise-aware* se añade a la entrada de las capas totalmente conectadas información adicional del ruido presente en una elocución en forma de

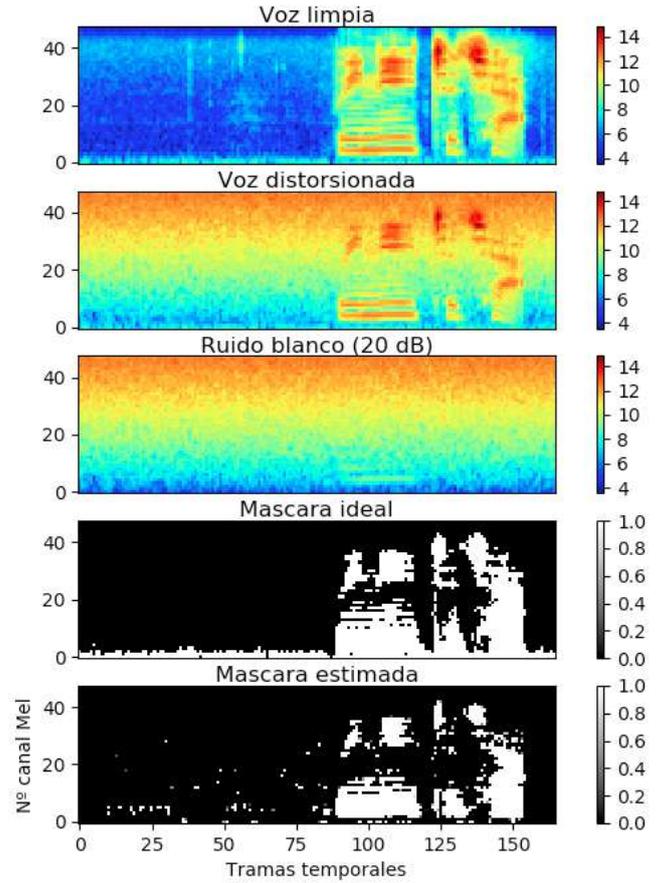


Fig. 4. Estimación de máscara con ruido blanco ( $SNR = 20\text{ dB}$ ). (a) Voz original sin ruido. (b) Voz distorsionada con ruido aditivo blanco a 20 dB de SNR. (c) Ruido blanco de 20 dB. (d) Máscara oráculo o ideal. (e) Máscara estimada.

vector con el fin de optimizar los parámetros del modelo, tal y como se muestra en la figura 3.

En este trabajo se propone el uso de máscaras de ruido como entrenamiento *noise-aware* para ofrecer a la red neuronal una información detallada sobre el ruido en cada bin tiempo-frecuencia. Una máscara define, para cada bin tiempo-frecuencia del espectro de la señal ruidosa, una puntuación que varía entre 0 y 1, indicativo de la probabilidad de que esa característica espectral esté contaminada por ruido. En la figura 4 se muestra una máscara de ruido oráculo (ideal) y una máscara estimada para una señal de voz contaminada con ruido blanco a una SNR de 20 dB. La máscara oráculo se obtiene dividiendo la potencia de la voz distorsionada entre la potencia real del ruido presente en la señal, mientras que la máscara estimada se calcula a partir de una estimación de la potencia del ruido. Esta estimación se obtiene a partir de una interpolación lineal con dos estimaciones de ruido independientes que se calculan promediando las  $T = 10$  primeras y últimas tramas de cada elocución. Una vez realizada la estimación del ruido, se estima la relación señal-ruido a partir de las estimaciones de ruido y de las características espectrales ruidosas originales:

$$\hat{SNR}(t, \lambda) = 10 * \log_{10} \left( \frac{Y(t, \lambda)}{\hat{N}(t, \lambda)} \right), \quad (3)$$

donde la tupla  $(t, \lambda)$  representa los bins tiempo-frecuencia,  $\mathbf{Y}$  son las FBANKs de la señal ruidosa, y  $\hat{\mathbf{N}}$  son las FBANKs estimadas del ruido. Finalmente, se emplea una función sigmoïdal centrada en 7 dB para comprimir la SNR estimada en el rango  $[0, 1]$ :

$$\widehat{\text{MÁSK}}(t, \lambda) = \frac{1}{1 + e^{7 - \widehat{\text{SNR}}(t, \lambda)}}. \quad (4)$$

#### D. Clasificador

Una vez que se obtiene el vector de identidad de suplantación de la elocución, se emplea un clasificador para decidir si la voz es genuina ó suplantada. En este trabajo se proponen 3 clasificadores:

1) *Análisis Discriminante Lineal*: El clasificador LDA asume que cada clase se puede modelar con una gaussiana multivariante

$$N(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)}, \quad (5)$$

donde  $\boldsymbol{\Sigma}_k$  y  $\boldsymbol{\mu}_k$  es la covarianza y la media de cada clase  $k$ , y  $p$  es la dimensión de los vectores de identidad. Además, el modelo LDA asume que cada clase comparte la misma covarianza, es decir,  $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$ ,  $\forall k$ . El objetivo de LDA es encontrar una transformación que maximiza la distancia entre clases y que minimice el esparcimiento dentro de cada clase. Esto se puede formular como un problema de diagonalización donde se diagonaliza la matriz  $\boldsymbol{\Sigma}_b \boldsymbol{\Sigma}$  ( $\boldsymbol{\Sigma}_b$  es la matriz de covarianza entre clases), de forma que la transformación se puede realizar a partir de los autovectores.

El clasificador LDA que se propone tiene  $K + 1$  clases que representan la voz genuina y los  $K$  ataques de suplantación presentes en el conjunto de entrenamiento. De esta forma, el LDA asigna un valor de confianza de voz genuina a cada elocución, el cual se usa para realizar una decisión binaria (voz genuina ó suplantada) en la fase de evaluación.

2) *Máquina de Soporte Vectorial*: Una máquina de soporte vectorial (SVM) separa los puntos definidos en un espacio de alta dimensión caracterizado por una función kernel. De esta manera, se obtiene primero una función binaria que describe la función de densidad de probabilidad donde se encuentran los datos de voz genuina. Esta función devuelve +1 en la pequeña región correspondiente a los datos de voz genuina, y -1 en el resto del espacio. Por tanto, la principal idea de SVM es estimar el hiperplano con el mayor margen de separación entre las dos clases.

3) *Máquina de soporte vectorial monoclasa*: Los clasificadores complejos pueden sobreajustar los datos de entrenamiento de voz suplantada. Con el fin de crear un sistema independiente del ataque de suplantación, también se va a evaluar un modelo derivado de SVM que solo se puede entrenar con datos de voz genuina. A este tipo de clasificador se le conoce como SVM monoclasa, usualmente empleado para encontrar datos anormales [13].

### III. FRAMEWORK DE EXPERIMENTACIÓN

Para evaluar el rendimiento del sistema de *anti-spoofing* propuesto, se va a emplear la base de datos ASVSpooof2015 [14], la cual contiene elocuciones sin distorsiones ruidosas

Conjunto	# Locutores		# Elocuciones	
	Hombre	Mujer	Genuina	Suplantada
Entrenamiento	10	15	3750	12,625
Validación	15	20	3497	49,875
Evaluación	20	26	9404	184,000

Tabla I

ESTRUCTURA DE LA BASE DE DATOS ASVspooof2015 [14].

con diferentes tipos de ataques de suplantación. Además, se emplea una versión ruidosa de dicha base de datos [5] con el fin de evaluar el rendimiento del sistema en entornos ruidosos y reverberantes. En esta sección se indican los detalles de las bases de datos y la metodología de entrenamiento.

#### A. Bases de datos

La base de datos ASVspooof2015 [14] define tres conjuntos (entrenamiento, validación y evaluación), donde cada uno contiene elocuciones genuinas y suplantadas. La estructura de estos tres conjuntos se muestra en la tabla I. Los ataques de suplantación se han generado mediante algoritmos de síntesis y conversión de voz. Se definen un total de 10 ataques de suplantación ( $S_1$  a  $S_{10}$ ): tres se implementan usando síntesis de voz ( $S_3, S_4, S_{10}$ ), y el resto se implementan con diferentes algoritmos de conversión de voz ( $S_1, S_2, S_5, S_6, S_7, S_8, S_9$ ). Los ataques conocidos, presentes en los tres conjuntos, son los ataques  $S_1$  a  $S_5$ , mientras que los ataques  $S_6$  a  $S_{10}$  son desconocidos, ya que solo están presentes en el conjunto de evaluación.

Para evaluar la robustez del sistema frente a distorsiones ruidosas y reverberantes, se ha empleado una versión ruidosa de la base de datos ASVspooof2015 [5]. Esta versión se ha generado artificialmente distorsionando las señales de voz originales con diferentes tipos de ruidos y con distintos niveles de SNR. En concreto, se han añadido 5 tipos de ruidos aditivos (ruido blanco, de gente hablando de fondo, de coche, exterior y cafetería) con 3 niveles de SNR (20, 10 y 0 dB). Además, se han añadido 3 escenarios reverberantes convolucionando las señales de voz limpias con 3 respuestas impulsivas cuyos valores T60 son: 0.3, 0.6 y 0.9 segundos. Por tanto, existen 18 condiciones ruidosas diferentes (15 ruidos aditivos y 3 condiciones reverberantes). Además, estas condiciones se han dividido en condiciones vistas y no vistas para realizar una evaluación más realista. Las condiciones vistas son los ruidos blanco, de gente hablando y exterior, y los 3 escenarios reverberantes, los cuales están presentes en los conjuntos de entrenamiento, validación y evaluación. Por otro lado, las condiciones no vistas contienen los ruidos de cafetería y coche, y están solo presentes en el conjunto de evaluación.

#### B. Entrenamiento

Las características espectrales de las señales de voz (FBANK, MFCC y CQCC) se calculan usando un banco de 48 filtros. Además, estas características se normalizan en media y varianza para entrenar el sistema de *anti-spoofing*.

En el escenario limpio, se emplea la base de datos [14] para entrenamiento y evaluación. En el escenario ruidoso, se emplea la base de datos [5], y en este caso, se aporta un vector de ruido a la salida de las capas convolucionales de la CNN para incrementar la robustez frente a las distorsiones.

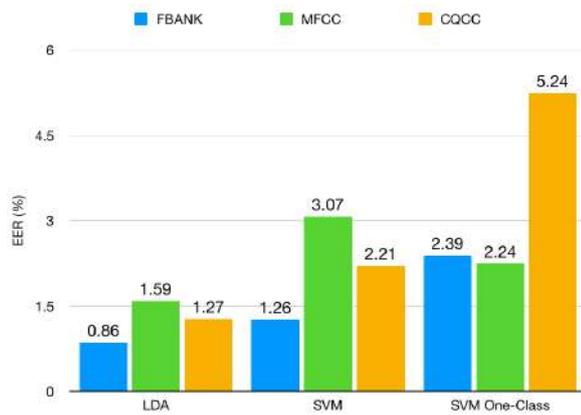


Fig. 5. Comparación de EERs (%) de ataques conocidos y desconocidos en entornos sin ruido sobre el conjunto de evaluación de ASVspooft2015 [14] con diferentes características espectrales (FBANK, MFCC y CQCC), y clasificadores (LDA, SVM y SVM monoclasa).

En ambos escenarios, la CNN y RNN se entrenaron usando el optimizador de Adam. Como existen  $K = 5$  ataques de suplantación en el conjunto de entrenamiento, la capa *softmax* de la CNN y RNN contiene  $K + 1 = 6$  neuronas (una por cada clase). Las dos capas totalmente conectadas de la CNN tienen 1024 neuronas, y la RNN tiene 1920 GRUs, que coincide con la longitud de los vectores de identidad de suplantación. Para prevenir el problema del sobreajuste, se ha inicializado  $prob[0] = 70\%$ . Además, se emplea un criterio de parada temprana con el fin de parar el entrenamiento cuando no se mejora la entropía cruzada tras 15 iteraciones. Todos los parámetros se han optimizado usando el conjunto de validación de la base de datos [14].

### C. Evaluación del rendimiento

Para medir el rendimiento del sistema se ha empleado la tasa EER (*Equal Error Rate*). Tal y como se describe en el plan de evaluación del reto ASVspooft 2015 [14], el EER se obtiene de forma independiente para cada ataque de suplantación, y después se calcula el EER promedio sobre todos los ataques.

## IV. RESULTADOS

### A. Comparación de características espectrales y clasificadores

En la figura 5 se muestran los resultados obtenidos en entornos sin ruido con las diferentes características espectrales (FBANK, MFCC y CQCC) y clasificadores (LDA, SVM y SVM monoclasa). El mejor rendimiento se obtiene con la combinación de características FBANK y clasificador LDA. En promedio, las FBANKs obtienen el mejor rendimiento independientemente del clasificador que se emplee, aunque las MFCCs obtienen mejor rendimiento con el clasificador SVM monoclasa.

En cuanto a los clasificadores, el LDA supera a los otros dos tanto en la detección de ataques conocidos como desconocidos. Además, el clasificador SVM binario supera al SVM monoclasa cuando se emplean las características FBANK y CQCC.

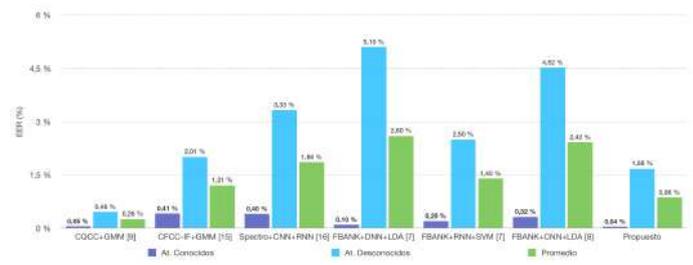


Fig. 6. Comparación de rendimiento en la detección de ataques conocidos y desconocidos en entornos sin ruido usando el conjunto de evaluación de ASVspooft2015 [14] en términos de EER (%).

De acuerdo a estos resultados, se propone para el sistema de *anti-spoofing* el empleo de características espectrales FBANK y un clasificador LDA para realizar la detección final (voz genuina ó suplantada).

### B. Evaluación en escenario sin distorsiones

En la figura 6 se muestra una comparación del rendimiento del sistema de *anti-spoofing* propuesto con los sistemas más relevantes de la literatura en escenarios sin ningún tipo de distorsión.

Los primeros dos sistemas CQCC + GMM [9] y CFCC-IF + GMM [15] emplean características espectrales CQCC y CFCC-IF con un clasificador GMM (*Gaussian Mixture Model*). Los otros 4 sistemas están basados en arquitecturas de redes neuronales. El sistema CQCC + GMM es el que obtiene el mejor rendimiento en promedio, aunque nuestro sistema (FBANK + CNN + RNN + LDA) lo supera en la detección de ataques conocidos.

Si se compara el sistema propuesto con el resto de sistemas basados en redes neuronales (Spectro + CNN + RNN [16], FBANK + DNN + LDA [7], FBANK + RNN + SVM [7] y FBANK + CNN + LDA [8]), nuestro sistema alcanza un mayor rendimiento tanto en la detección de ataques conocidos como desconocidos. Además, nuestra propuesta supera en la detección de casi todos los ataques al sistema CFCC-IF [15], alcanzando un EER 0.35% menor en el promedio de todos los ataques.

### C. Evaluación en entornos ruidosos y reverberantes

En esta sección se compara el sistema de *anti-spoofing* propuesto sobre el conjunto de evaluación de la versión ruidosa de la base de datos ASVspooft 2015 [5] con el mejor sistema de *anti-spoofing* en entornos sin distorsiones (CQCC + GMM), y con el sistema que emplea una CNN y la técnica de estimación promedio del ruido propuesto en [8] (CNN + NAT).

En la figura 7 se muestra el rendimiento de los tres sistemas cuando se evalúan en un escenario sin ningún tipo de distorsión, y cuando se entrenan con señales de voz ruidosas. A pesar de que en la sección anterior se demuestra que el sistema CQCC + GMM es el que obtiene mayor rendimiento en un escenario sin ruido, el sistema propuesto le supera en un 0.03 % en promedio cuando el entrenamiento se realiza con datos ruidosos. Esto se debe a la capacidad que tiene el sistema propuesto para generalizar, siendo capaz de mejorar

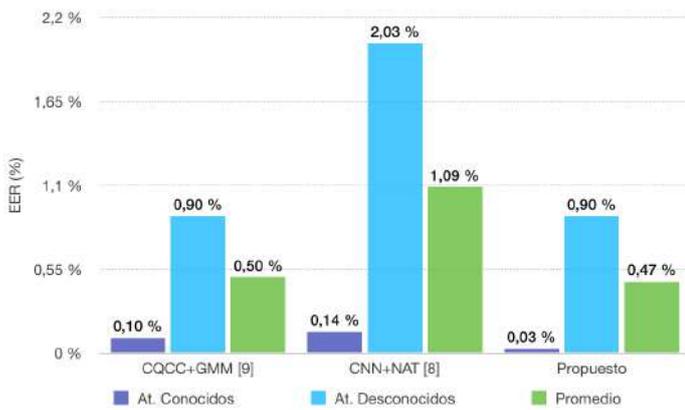


Fig. 7. Evaluación de sistemas de *anti-spoofing* en entornos sin distorsiones ruidosas usando datos de entrenamiento ruidosos en términos de EER (%).

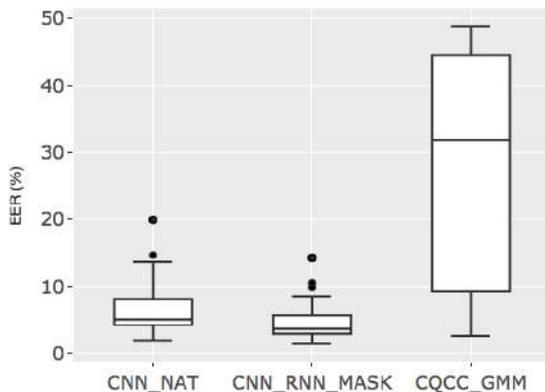


Fig. 8. Comparación de sistemas de *anti-spoofing* en escenario con distorsiones ruidosas y reverberantes en términos de EER (%). El sistema propuesto (CNN + RNN + MASK) es el que alcanza un mayor rendimiento.

al resto de sistemas cuando se evalúan tanto en un escenario limpio como con distorsiones.

Por otro lado, en la figura 8 se compara el rendimiento de los tres sistemas en escenarios con distorsiones ruidosas. El sistema propuesto mejora en un 1.6 % y 1.7 % al CNN + NAT en los escenarios de distorsiones vistas y no vistas durante el entrenamiento, respectivamente. Además, es destacable cómo disminuye el rendimiento del sistema CQCC + GMM en un escenario con distorsiones ruidosas, ya que el sistema propuesto le supera en un 26.6 % y 18.5 % en los escenarios vistos y no vistos en la fase de entrenamiento, respectivamente.

Por tanto, la técnica propuesta de máscaras de ruido supera el rendimiento de la técnica de estimación promedio del ruido propuesta en [8], debido a que las máscaras de ruido ofrecen información sobre la probabilidad de presencia de ruido en cada bin tiempo-frecuencia, a diferencia de la otra técnica que solo ofrece información del promedio del ruido presente en la señal de voz.

## V. CONCLUSIONES

En este trabajo se han realizado tres contribuciones principales al estado de la técnica de detección de ataques de suplantación de voz: (i) estudio de las mejores características espectrales y clasificadores para sistemas basados en redes neuronales, (ii) obtención de un único vector de características

de suplantación por cada elocución usando una RNN, y (iii) uso de máscaras de ruido para alcanzar el mejor rendimiento en escenarios desfavorables de ruido y reverberación.

Este trabajo ha supuesto la concesión de una beca FPU para realizar una tesis doctoral con título "Reconocimiento Robusto de Locutores Basado en Redes Neuronales Profundas", donde se pretende abordar completamente el problema de verificación de locutores para su aplicación en sistemas de autenticación biométrica basada en voz.

## CONTRIBUCIONES

Los resultados de este trabajo han sido publicados en 2 congresos internacionales: Iberspeech 2018 [17] e Interspeech 2018 [18]. Este último se trata del congreso anual más importante de tecnologías del habla a nivel internacional. Además, actualmente se está trabajando en un artículo de revista usando como base las técnicas de este trabajo.

## REFERENCIAS

- [1] M. Khitrov, "Taking passwords: voice biometrics for data access and security," *Biometric Technology Today*, vol. 2013, no. 2, pp. 9–11, 2013.
- [2] Roberto Togneri and Daniel Pallela, "An Overview of Speaker Identification: Accuracy and Robustness Issues," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 74–99, 2015.
- [3] Z. Wu, N. Evans, J. Yamagishi, T. Kinnunen, et al. "ASVspoof: The Automatic Speaker Verification Spoofing and Countermeasures Challenge," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 4, pp. 588–604, 2017.
- [4] Z. Wu, N. Evans, T. Kinnunen, et al., "Spoofing and countermeasures for speaker verification: A survey," *Speech Communication*, vol. 66, no. 4, pp. 788–798, 2015.
- [5] X. Tian, Z. Wu, X. Xiao, et al., "An investigation of spoofing speech detection under additive noise and reverberant conditions," en *Proc. Interspeech*, 2016.
- [6] M. P. Cooke, P. Green, L. Josifovski, and A. Vizinho, "Robust automatic speech recognition with missing and unreliable acoustic data," *Speech Communication*, vol. 34, pp. 267–285, 2001.
- [7] Yanmin Qian, Nanxin Chen, Kai Yu, "Deep Features for automatic spoofing detection," *Speech Communication*, vol. 85, no. 1, pp. 43–53, 2016.
- [8] Yanmin Qian, Nanxin Chen, Heinrich Dinkel, Zhizheng Wu, "Deep Feature Engineering for Noise Robust Spoofing Detection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 10, pp. 1942–1955, 2017.
- [9] M. Todisco, H. Delgado, and N. Evans, "A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients," *Proc. Odyssey*, 2016.
- [10] S. Rennie, V. Goel, and S. Thomas, "Annealed dropout training of deep networks," in *Proc. Spoken Language Technology Workshop*, 2014.
- [11] Kyunghyun Cho, et al., "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," *Proc. EMNLP*, 2014.
- [12] J. C. Segura, M. C. Benítez, A. de la Torre, S. Dupont, A. J. Rubio, "VTS residual noise compensation," *Proc. ICASSP*, 2002.
- [13] Scholkopf, B., Williamson, R. C., Smola, A. J., et al., "Support vector method for novelty detection," in *Proc. NIPS*, 2000, pp. 582–588.
- [14] Wu, Z., Kinnunen, T., et al., "ASVspoof 2015: The First Automatic Speaker Verification Spoofing and Countermeasures Challenge," *Proc. Interspeech*, 2015.
- [15] T. B. Patel and H. A. Patil, "Combining evidences from mel cepstral, cochlear filter cepstral and instantaneous frequency features for detection of natural vs. spoofed speech," *Proc. Interspeech*, 2015.
- [16] C. Zhang, C. Yu and J. H. L. Hansen, "An Investigation of Deep-Learning Frameworks for Speaker Verification Antispoofing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 4, pp. 684–694, 2017.
- [17] A. Gomez-Alanis, A. M. Peinado, J. A. Gonzalez and A. Gomez, "Performance evaluation of front- and back-end techniques for ASV spoofing detection systems based on deep features," *Proc. Iberspeech*, 2018.
- [18] A. Gomez-Alanis, A. M. Peinado, J. A. Gonzalez and A. Gomez, "A Deep Identity Representation for Noise Robust Spoofing Detection," *Proc. Interspeech*, 2018.

# Índice de autores

Alberto Abellán Galera.....	127, 133	Óscar Adamuz Hinojosa .....	111
Cristian Almohalla Gómez .....	169	Juan Carlos Angulo Santos .....	61
Pedro José Arias Perandrés.....	163	Gabriel Ávila Muñoz .....	21
Fernando Bayo Martínez .....	201	Miguel Bolívar Lupiáñez .....	217
José Carreto Morales .....	175	Lorena Chinchilla Romero .....	105
Natalia Chinchilla Romero .....	43	Francisco José Cuberos Muñoz .....	9
Félix Delgado Ferro .....	87	Laura Gámiz Pérez.....	241
Francisco García Espigares .....	195	José Manuel García Giménez .....	3
Juan Miguel García Martínez.....	157	Pablo Jesús García Martínez .....	253
Alejandro Gómez Alanís .....	271	José Alberto Gómez López .....	223
Miriam González Atienza.....	265	Ángel Guzmán Martínez .....	37
Bilal Hammu Mohamed .....	183	Pablo Helio Zapata .....	93
Joaquín Lara Sarabia .....	55	Irene Herrera López .....	15
Pablo Manresa Nebot.....	207	Juan Manuel Martín Doñas.....	247
José Miguel Mateos Ramos .....	229	Pablo Merino Ávila.....	145
Mario Molina Molina.....	235	Javier Montijano Valdivia .....	31
Segundo Moreno Torres.....	151	José Miguel Noguera Comino .....	139
José Antonio Ordóñez Lucena .....	99	Sergio Ortiz Ruiz .....	211
Ángel Palomares Caballero .....	77	David Pareja Martos .....	259
Francisco Jesús Quero de la Rosa.....	83, 117	Beatriz Quesada Tobar .....	123
Alejandro Ramírez Arroyo .....	67	Pablo Rodríguez Martín.....	49
Cleofás Segura Gómez .....	189	César Senés Romo .....	27
Miguel Ángel Vaquero Blasco.....	71		



# Índice de tutores

Pablo Ameigeiras Gutiérrez.....	M <sup>a</sup> Carmen Benítez Ortúzar ..... 241
..... 55, 99, 105, 111,	
José Camacho Páez.....3, 9, 15, 145	Luz García Martínez ..... 241
Francisco Javier García Ruiz..... 183	Pedro García Teodoro..... 123
Ángel M. Gómez García .....	José Andrés González López.....
..... 247, 253, 259	..... 265, 271
Juan Manuel Górriz Sáez .....	Miguel Ángel López Gordo .....
..... 235	..... 21
Gabriel Maciá Fernández .....	Roberto Magán Carrión .....
..... 127, 133, 139	..... 133
Antonio M. Mora García..... 145, 151,	Pablo Muñoz Luengo.....
	..... 83, 93, 105, 117
Jorge Navarro Ortiz .....	Pablo Padilla de la Torre .....
..... 27, 37, 43, 61, 87, 195	..... 67, 71, 77, 183, 189
Antonio M. Peinado Herreros .....	Jonathan Prados Garzón..... 31, 157
..... 247, 253, 259, 265, 271	
Javier Ramírez Pérez de Hínestrosa ...	Juan José Ramos Muñoz .....
..... 217, 235	.... 31, 55, 99, 111, 157, 163, 169, 175
Diego Salas González .....	José Carlos Segura Luna .....
..... 201, 217, 229, 235	..... 195
Sandra Sendra Compte .....	Fermín Segovia Román.....
..... 49	..... 201, 207, 211, 223
Juan Francisco Valenzuela Valdés.....	
..... 67, 71, 77, 83, 93, 117, 189,	





