



UNIVERSIDAD  
DE GRANADA

TRABAJO FIN DE GRADO  
INGENIERÍA DE TECNOLOGÍAS DE TELECOMUNICACIÓN

# Integración de 5G y TSN en redes privadas industriales

---

Network Slicing en redes de transporte  
TSN asíncrono

## Autor

Julia Caleyá Sánchez

## Directores

Pablo José Ameigeiras Gutiérrez  
Jonathan Prados Garzón



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE  
TELECOMUNICACIÓN

—  
Granada, 6 de septiembre de 2021

# Integración de 5G y TSN en redes privadas industriales

---

**Network Slicing en redes de transporte**  
**TSN asíncrono**

## **Autor**

Julia Caleyá Sánchez

## **Directores**

Pablo José Ameigeiras Gutiérrez

Jonathan Prados Garzón



# **Integración de 5G y TSN en redes privadas industriales: *Network Slicing* en redes de transporte TSN asíncrona**

Julia Caleyá Sánchez

**Palabras clave:** red 5G, URLLC, red TSN, Industria 4.0, TSN asíncrono, *backhaul*, 5G Core, *network slicing*, *slices*, *QoS*, CNC, TSN AF, simulador.

## **Resumen**

La revolución industrial ha provocado la aparición de nuevos tipos de servicios que requieren las calidades de servicio propias de las comunicaciones cableadas, pero con la posibilidad de movilidad de los dispositivos de la red. Una de las finalidades de la creación de las redes inalámbricas de quinta generación (5G) es satisfacer estos nuevos servicios. A pesar de su utilización en la industria, un sistema 5G no es capaz de cumplir con la alta fiabilidad y el bajo retardo exigidos. Sin embargo, la tecnología cableada de capa 2 *Time-Sensitive Network (TSN)* si proporciona estos requerimientos de fiabilidad y latencia.

Por tanto, una solución al problema de la Industria 4.0 es la integración de las redes 5G y TSN, tema tratado por numerosas investigaciones. Existe ya un modelo estandarizado por el 3rd *Generation Partnership Project (3GPP)* para su integración, donde el sistema 5G se convierte en un puente TSN. Este modelo no contempla la posibilidad de desarrollar la funcionalidad de *network slicing* de los sistemas 5G, siendo una característica fundamental para incrementar la utilización de la red. Además, considera que la industria previamente tiene implementada una red TSN, pero puede suceder que este desarrollada por una red 5G.

Por lo que en este proyecto se trata de encontrar la solución para soportar *network slicing* en las redes 5G industriales, donde se utiliza una red TSN asíncrona para la implementación de la red de transporte. Con la integración de TSN asíncrono en la red de transporte se origina el problema de asignación de flujos a las distintas *slices* de la red. Este problema se resuelve con el desarrollo de un optimizador que minimiza la probabilidad de rechazo global de la red, gracias a las herramientas *CVX* y *MOSEK* en *Matlab*.

Finalmente, se implementa el optimizador en un simulador de una red TSN asíncrona con la funcionalidad de *network slicing*. Se lleva a cabo diversos experimentos para estudiar la granularidad en términos de número de *slices* adecuada para dar servicio a los diferentes servicios de la Industria 4.0, cuyas características también se ha definido en este proyecto.



# Integration of 5G and TSN in industrial private networks: *Network Slicing* in asynchronous TSN transport networks

Julia Caleyá Sánchez

**Keywords:** 5G network, URLLC, TSN network, Industry 4.0, asynchronous TSN, backhaul, network slicing, slices, QoS, 5G Core, CNC, TSN AF, simulator.

## **Abstract**

The industrial revolution has caused the appearance of new types of services that require the qualities of service suitable to the wiring communication, but with the possibility of movement of the network devices. One of the purposes of the creation of the wireless network of the fifth generation (5G) is to satisfy these new services. In spite of its use in the industry, a 5G system is not able to fulfill with the high reliability and the low needed delay. However, the wiring technology of layer 2 *Time-Sensitive Network* (TSN) can supply these requirements of reliability and latency.

Therefore, one solution to the problem of the 4.0 Industry is the integration of the 5G and TSN networks, this theme is dealt by numerous investigations. There is already a standardized model by the 3rd Generation Partnership Project (3GPP) for its integration, where the 5G system is converted into a TSN bridge. This model does not contemplate the possibility to develop the functionality of network slicing of the 5G systems, being one of the fundamental characteristic to increase the use of the network. Besides, it is considered that the industry has previously implemented a TSN network, but it is also possible that it can be developed by a 5G network.

So, this project wants to find a solution to support network slicing in the 5G industrial networks, where it is used an asynchronous TSN network to the implementation of the transport network. The integration of asynchronous TSN in the transport network causes the problem of the assignment of flows to the different slices of the network. This problem is solved with the development of an optimizer that minimize the probability of global rejection in the network, thanks to the *CVX* and *MOSEK* tools in *Matlab*.

Finally, the optimizer is implemented in a simulator of an asynchronous TSN network with the functionality of network slicing. Some experiments are carried out in order to study the granularity in terms of numbers of correct slices to give service to the different services of the 4.0 Industry, whose characteristics have also been defined in this project.



---

D. **Pablo José Ameigeiras Gutiérrez**, Profesor del Área de Telemática del Departamento Teoría de la señal, Telemática y Comunicaciones de la Universidad de Granada.

D. **Jonathan Prados Garzón**, Profesor del Área de Telemática del Departamento Teoría de la señal, Telemática y Comunicaciones de la Universidad de Granada.

### **Informan:**

Que el presente trabajo, titulado *Integración de 5G y TSN en redes industriales privadas, Network Slicing en redes de transporte TSN asíncronas*, ha sido realizado bajo su supervisión por **Julia Caleyá Sánchez**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a 6 de septiembre de 2021.

### **Los directores:**

**Pablo José Ameigeiras Gutiérrez**

**Jonathan Prados Garzón**



# Agradecimientos

Tras la finalización de este proyecto que pone punto final a los cuatro años de carrera me gustaría agradecer a todas aquellas personas que han hecho posible esta etapa, que sin ellos hubiera sido más cuesta arriba.

En primer lugar, a mi familia, por ser mi apoyo incondicional en todo momento desde el primer momento que decidí realizar esta carrera, donde la distancia no ha sido un inconveniente, sino que ha permitido fortalecer las relaciones existentes. En especial, a mis padres, Francisco y Rosario, por ofrecerme la oportunidad y todas las facilidades del mundo para estudiar la carrera que quería, sin importar nada con tal de ofrecerme lo mejor para mí y para mis hermanos. Además, de enseñarme los mejores legados y valores que cualquier hijo quisiera tener: el trabajo, la formación y la honradez, entre muchos otros. A mi hermana mayor, Alejandra, que siempre ha sido mi ejemplo a seguir, y a mi hermano menor, Paco, por enseñarme a cuestionarme y a no conformarme con nada en esta vida. Gracias a vosotros me habéis convertido en la persona que soy hoy en día.

Agradezco también a todos aquellos profesores, tanto universitarios como aquellos de años anteriores, que me han formado e instruido hasta ahora, y han despertado en mis inquietudes que no existían. Llegar hasta aquí no hubiera sido posible sin vosotros.

A mis tutores, Pablo y Jonathan, por su total disponibilidad siempre que lo he necesitado y por confiar en mí para realizar este proyecto. Especialmente, por mostrarme el mundo de la investigación. Además de vuestra comprensión y paciencia, y por el tutelaje no solo académico, sino también personal y profesional.

Por supuesto, dar las gracias a todos mis amigos, a mis compañeros de carrera, a mis amistades de Fuente del Maestre y sobre todo, a mis compañeras de piso y a mis amigos de Granada. Sin la diversión, la comprensión y la confianza que me aportáis este proyecto habría sido mucho más difícil.

Finalmente, agradecer a todo profesional que ha dado su vida en este periodo de pandemia poder volver a la normalidad lo más pronto posible.

Gracias a todos.



# Índice general

1. Introducción .....	22
1.1 Industria 4.0.....	23
1.2 Tecnologías habilitadoras.....	25
1.2.1 Redes de banda ancha de Quinta Generación (5G).....	25
1.2.2 Time-Sensitive Networking (TSN) .....	27
1.3 Motivación .....	29
1.4 Objetivos .....	30
1.5 Organización .....	31
2. Estado del arte .....	35
2.1 5G .....	35
2.2 TSN .....	39
2.2.1 TSN síncrono.....	46
2.2.2 TSN asíncrono.....	51
2.2.3 Ventajas e inconvenientes de TSN síncrono y asíncrono.....	54
2.3 Integración de TSN y 5G .....	54
2.3.1 Arquitectura de integración de TSN y 5G según el 3GPP .....	54
2.3.2 Proceso de interacción entre el plano de control de 5GS y el CNC de los puentes TSN .....	57
2.3.3 Proceso de sincronización temporal.....	60
2.3.4 Dominios de tiempo .....	61
2.3.5 Alta fiabilidad en redes TSN .....	62
3. TSN para la red de transporte de una red 5G .....	64
3.1 Network Slicing en 5GS.....	65
3.2 Diseño .....	68
4. Evaluación Experimental .....	73
4.1 Marco de evaluación .....	73
4.2 Herramienta de evaluación.....	77
4.2.1 Simulador .....	77
4.2.2 CVX .....	79

4.2.3 MOSEK.....	80
4.3 Escenario de evaluación.....	81
4.4 Resultados .....	86
5. Conclusiones y vías futuras.....	91
6. Bibliografía .....	94
Apéndice A. Cálculo de la probabilidad de rechazo de flujo al 10% .....	98

# Índice de figuras

Figura 1: Transformación de la industria [29].....	24
Figura 2: Evolución de las versiones publicadas por el 3GPP [2] .....	26
Figura 3:Características TSN [15].....	28
Figura 4: Temporización del proyecto. ....	33
Figura 5: Evolución de la red 5G. [19].....	37
Figura 6: Modelo completamente distribuido TSN.....	40
Figura 7: Modelo de red centralizado y usuario distribuido TSN .....	40
Figura 8:Modelo completamente centralizado TSN [29].....	41
Figura 9: Intercambio de información LLDP [5] .....	42
Figura 10: Mapeo entre los PCP y las clases de tráfico con IEEE 802.1Q [29] .....	44
Figura 11: Funcionamiento del mecanismo FRER [29].....	45
Figura 12: Ejemplo de un dominio gPTP donde se comparte un reloj según IEEE 802.1AS [16] .....	46
Figura 13: Intercambio de mensajes para calcular el retardo de propagación .....	47
Figura 14: Varios dominios de sincronización temporal dentro de una fábrica [29] .....	49
Figura 15: División del tiempo en intervalos temporales [29] .....	49
Figura 16: Ejemplo del funcionamiento del mecanismo de control de puertas [29].....	50
Figura 17: Procesamiento de tramas considerando IEEE 802.1Qci, IEEE 802.1Qbv, IEEE 802.1Qbu y IEEE 802.3br [4] .....	50
Figura 18: Arquitectura y funcionamiento de UBS [17].....	52
Figura 19: Arquitectura y funcionamiento del scheduler ATS [17].....	53
Figura 20: Arquitectura de integración TSN-5G del plano de usuario [15].....	55
Figura 21: Arquitectura de integración TSN-5G con más de un puente virtual TSN [29].....	56
Figura 22: Arquitectura de integración TSN y 5G .....	57
Figura 23: Interacción en el plano de control entre TSN y 5G [29].....	58
Figura 24: Mapeo de QoS entre tráfico TSN y flujos QoS 5G [29].....	59
Figura 25: Sincronización temporal de la integración entre TSN y 5G .....	61
Figura 26: Sincronización temporal de la arquitectura TSN-5G con varios dominios temporales .....	62
Figura 27: Posibles soportes para las comunicaciones ultra confiables [29] .....	62
Figura 28: 5G network slices ejecutándose en una misma infraestructura física. [11] .....	65
Figura 29: Principio de recursividad con InPs y tenants virtualizados. [11].....	67
Figura 30: Orquestación y gestión de los slices de red. [4].....	68
Figura 31: Arquitectura de la red de transporte backhaul. [33].....	71
Figura 32: Arquitectura de gestión y orquestación de la red de transporte. [17] .....	71
Figura 33: Topología de red industrial. [17] .....	81
Figura 34: Representación de la tasa agregada en función de la probabilidad de rechazo para las slices de tipo eMBB según su topología y el número de slices.....	88
Figura 35: Tasa agregada (Mbps) en función de la probabilidad de rechazo del 10% .....	98

# Índice de Tablas

Tabla 1: Mapeo de los 5QI a las características de 5G QoS .....	39
Tabla 2: Valores PCP según la criticidad y el tipo de tráfico [29] .....	43
Tabla 3: Comparación entre redes TSN síncronas y asíncronas [7].....	54
Tabla 4: Características de las slices de automatización de fábrica. ....	83
Tabla 5: Características de las slices de automatización de procesos. ....	84
Tabla 6: Características de las slices de HMI.....	85
Tabla 7: Características de las slices del ámbito de supervisión y mantenimiento .....	85
Tabla 8: Características de las slices del ámbito eMBB .....	86
Tabla 9: Tasa agregada global para una probabilidad de rechazo del 10% según la topología. .	87
Tabla 10: Tasa agregada global para una probabilidad de rechazo del 10% según la combinación de slices .....	89



# Capítulo 1

## Introducción

Actualmente cada vez es más común el conexionado de dispositivos y aparatos a Internet para poder controlarlos y comprobar su correcto funcionamiento. Estos dispositivos tienen diferentes ámbitos de actuación desde electrodomésticos del propio hogar, robots y maquinaria utilizadas en fábricas y hasta en grandes almacenes, refinerías o en las propias ciudades. Una de las principales características que desean estos dispositivos es disponer de una gran movilidad, por lo que necesitan estar conectados a redes inalámbricas (*wireless*) en lugar de a redes cableadas tradicionales. Además, otra gran ventaja de utilizar redes *wireless* es su facilidad de despliegue que permite la reducción del coste de instalación, ya que no es necesario un cable que conecte con cada dispositivo. Sin embargo, las redes inalámbricas presentan unas características del medio que provocan pérdidas de potencia en las ondas electromagnéticas que se utilizan para la comunicación entre los distintos nodos de la red. Este hecho provoca que el tamaño de las redes *wireless* no sea tan extenso como en las cableadas. Aunque en determinados sectores industriales, que requieren las características proporcionadas por las redes *wireless*, presentan una infraestructura extensa. Por tanto, se necesita de nuevos protocolos que subsanen dicho problema.

Las redes inalámbricas de área extensa (*Wireless Wide Area Network*, WWANs) permiten dar cobertura a dispositivos *wireless* en áreas geográficas de gran extensión. Estas redes presentan equipos específicos, denominados puntos de accesos, que posibilitan la comunicación entre los equipos inalámbricos, denominados estaciones bases y otras células o redes. Por consiguiente, este modo de funcionamiento se denomina de infraestructura, donde se requiere que un los equipos y los sistemas estén desplegados previamente. Este tipo de tecnología ha avanzado enormemente a lo largo de los años, ya que desde su inicio permitió la comunicación entre las personas que se encontraban a grandes distancias entre ellas. Además, con el surgimiento de Internet al mundo, creado en 1983 por el departamento de Defensa de Estados Unidos, se consiguió la creación de nuevos servicios y dispositivos para tener acceso a ellos, como las consultas remotas de archivos. El desarrollo de estos nuevos servicios también se incluyó en las redes celulares, donde se crearon nuevos terminales que además de las llamadas de voz normales también se permitía la comunicación a través de datos, con mensajería instantánea, las búsquedas por Internet y estos últimos años, las comunicaciones a través de redes sociales. Esto provocó el surgimiento de nuevos protocolos de comunicaciones celulares que permitían incrementar la tasa de información que se transmitiese en cada envío de trama y los tamaños de los paquetes,

reduciendo los tiempos de latencia. Además, se consiguió aumentar el número de dispositivos que podía acceder a cada estación base o punto de acceso para que todos los usuarios que lo desearan pudieran acceder. Por consiguiente, el principal objetivo de esta continua evolución es facilitar la vida de los usuarios.

Recientemente con la inclusión de la nueva tecnología celular de banda ancha (5G) en el mundo se han desarrollado estudios [13] donde consideran que se generará grandes cantidades de dinero, principalmente en el entorno de la fabricación.

El uso de las redes inalámbricas celulares se ha expandido llegando también a las fábricas, sufriendo, por tanto, grandes cambios en su modo de funcionamiento. De este modo, se crean nuevas oportunidades de negocios y funcionalidades como la utilización de robots autónomos que permiten el ahorro de tiempo a lo hora de ejecutar las tareas, así como una mayor flexibilidad. El control de caminos también es otra funcionalidad añadida para llevar a cabo rutas sin tripulación para el reparto o el transporte de materiales, por ejemplo. Esto se puede conseguir gracias a toda la información que obtiene de la conexión con otros automóviles y las propias señales de tráfico de las ciudades y las carreteras. También puede permitir el uso de drones en pequeños comercios para el reparto o en los almacenes para organizar o transportar los materiales. Por tanto, no solo afecta al sector industrial, sino que agrupa a varios dominios de uso. De este modo, se puede llevar a cabo la automatización de electrodomésticos para realizar la compra una vez que se ha finalizado un determinado producto, o llamar al técnico cuando se produce un fallo técnico.

Por consiguiente, gracias a la próxima incorporación de las redes celulares, con su última versión (actualmente 5G) a la industria se permite crea nuevas funcionalidades que cambiarán la forma de trabajo y la creación de los nuevos productos según las características deseadas de los individuos y la gestión de los equipos que intervienen en su producción. A continuación, se explica en más detalle las principales características de la Industria 4.0 y las tecnologías que permiten su correcto funcionamiento. También se indican las motivaciones y los objetivos de este proyecto, además de la planificación que se ha seguido para su desarrollo.

### **1.1 Industria 4.0**

La industria ha sufrido grandes cambios a lo largo de la historia desde la primera revolución industrial que se produjo en 1760, que introdujo la utilización de equipos de producción mecánicos. Estos equipos de producción mecánicos se impulsaban por el agua y la energía del vapor, como por ejemplo la máquina de vapor. Con la llegada de la segunda revolución, se incorporó la producción en cadena y el uso de energía eléctrica en lugar de utilizar el vapor. En esta etapa se creó por primera vez la cinta transportadora y, por tanto, la fabricación en cadena. La tercera revolución industrial provocó el uso de la electrónica y la informática para impulsar la producción automatizada y la digitalización de la información, con el fin de facilitar dichos datos a los empresarios y el control de la producción a través de robots, que simplifican determinados procesos de fabricación. Esta automatización se realizaba con redes construidas con tecnologías de bus, que son muy costosas y difíciles de interoperar entre ellas, debido al amplio abanico disponible de estas tecnologías, cada una de ellas destinadas a proporcionar conectividad a un determinado servicio. Por último, la cuarta revolución industrial introduce la conectividad sensible al tiempo entre las diferentes fábricas para ser más eficientes y productivas, creando sistemas autónomos con gran escalabilidad, escasos fallos y menor tiempo de espera entre órdenes para conseguir la mayor interacción posible.

Por tanto, con esta revolución no solamente se ven afectados los procesos de fabricación, sino que puede afectar a toda la industria en general y a la propia sociedad. Ya que con la introducción de la conectividad inalámbrica se puede llevar a cabo la comunicación entre los diferentes agentes que intervienen en el proceso de producción, como son los proveedores, los clientes, los inversores, etc. Esto provoca un incremento en el intercambio de datos que se produce entre los múltiples sistemas y los propios participantes del proceso de fabricación. Asimismo, las comunicaciones entre los componentes que conforman la fábrica y que están conectados inalámbricamente pueden ajustarse y transmitir datos de configuración en tiempo real para evitar posibles fallos o problemas en la producción. De esta manera, las máquinas deben ser suficientemente inteligentes para poder analizar y diagnosticar sus propios problemas sin intervención humana. Pero la principal característica que definen a la Industria 4.0 es el Internet de las cosas (*Internet of Things, IoT*), es decir, como se ha comentado anteriormente, se aumenta la automatización de todos los procesos a través de la mejora de las comunicaciones entre los nodos y el autocontrol.

La Industria 4.0 presenta una serie de características necesarias para llevar a cabo los nuevos servicios que se pueden implementar. Algunas de estas características son una baja latencia, una alta confiabilidad y densidad de conexión, además de permitir la movilidad de las estaciones base. Es decir, una baja latencia es necesaria para que los paquetes lleguen inmediatamente desde un extremo a otro para evitar problemas por ejemplo en la configuración de una máquina en un determinado momento. La alta confiabilidad es importante, ya que los nodos deben saber que los paquetes recibidos no han sufrido ningún fallo o han sido modificados en el proceso de transmisión. Ya que los servicios que están ejecutando estos nodos no es posible la retransmisión de los datos continuamente, debido a que provocaría problemas en el correcto funcionamiento del servicio que están ofreciendo. Y por último una alta densidad de conexión, es decir que permita dar cobertura a un número elevado de clientes con calidades de servicio estrictas.

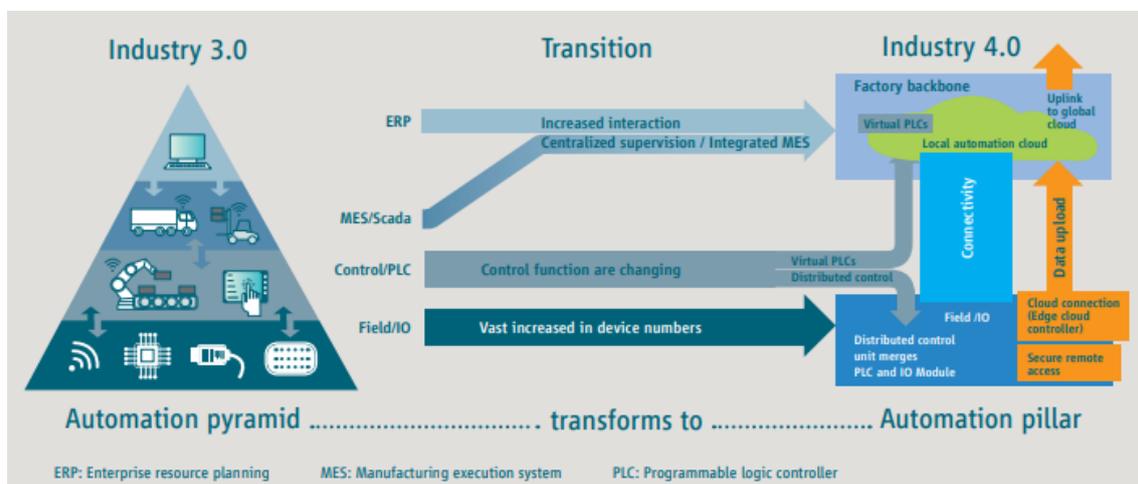


Figura 1: Transformación de la industria [29]

Este conjunto de requisitos son las que se ofrecen en una red cableada, pero que se debe conseguir en nuevas tecnologías que permitan la movilidad de los clientes que conforman la red. Las tecnologías existentes no garantizan que una determinada tarea se ejecute en el momento deseado, ya que, con Ethernet, por ejemplo, no se proporcionan los requisitos de determinismo que necesita la Industria 4.0 debido al uso de *Carrier Sense Multiple Access (CSMA)* como mecanismo de prevención de fallos en la transmisión.

## **1.2 Tecnologías habilitadoras**

Los requisitos necesarios para la implementación de las Industrias 4.0 no son posibles con las tecnologías actuales, debido a que las anteriores generaciones de redes celulares no garantizaban características tan importantes como una baja latencia o una alta fiabilidad. Actualmente, el desarrollo de las nuevas fábricas es posible gracias a la creación de las redes de banda ancha de quinta generación (5G). Esta tecnología ofrece una mayor flexibilidad con un menor coste y una baja latencia, es decir, los plazos de entrega de los paquetes relativamente más cortos que generaciones anteriores. Sin embargo, las redes industriales modernas presentan muchos casos de uso donde la sincronización con el tiempo, como sucede en el control del movimiento de bucle cerrado, así como al uso de la nube es un aspecto importante. Por lo tanto, también se debe tener en cuenta que las redes industriales deben ser tolerantes a fallos y con una sincronización de tiempo precisa. Además, para llevar a cabo una correcta sincronización de tiempo es necesario la programación de radio de bajo consumo y una coordinación distribuida dentro de la red.

Por consiguiente, uno de los principales problemas que presenta el 5G en las zonas que se está desplegando es su incapacidad de dar tasas de retardo tan bajas y con tanta confiabilidad como las requeridas por la Industria 4.0, no solamente en la parte radio sino también en las tecnologías subyacentes de las comunicaciones extremo a extremo. Por lo que se necesita incorporar otras tecnologías que lo solucionen. En este caso, se integra una colección de estándares definidos en el IEEE 802.1Q denominados redes de tiempo sensible (*Time Sensitive Networking, TSN*), que permiten reducir el tiempo de retardo, así como la sincronización de tiempo entre los dispositivos que conforman la red.

### **1.2.1 Redes de banda ancha de Quinta Generación (5G)**

La telefonía móvil ha sufrido grandes cambios en los últimos años debido a la evolución de los servicios con nuevas prestaciones y al aumento de los usuarios, que provoca la gran variabilidad en la creación de las nuevas generaciones. Los diferentes sistemas de telefonía con un marco común se agrupan formando nuevas generaciones. De esta forma surge la generación cero, donde se produce las primeras comunicaciones públicas de forma comercial a través del radioteléfono móvil, que requerían formar parte de un vehículo. La primera generación tenía el propósito de llevar la telefonía inalámbrica a todos los usuarios con el uso de tecnologías analógicas. Con la llegada del 2G se busca la interconexión entre las redes y el uso de la tecnología digital, con la aparición del estándar GSM para usar de forma eficiente el espectro de radiofrecuencias y dar por primera vez un servicio de mensajería de texto e intercambio de imágenes. El principal objetivo de la tercera generación era la combinación de la conmutación de paquetes y de circuitos para alcanzar mayores velocidades de servicios de datos. Aunque no se consiguieron altas velocidades (superiores a 100 Mbps de bajada y 50 Mbps de subida) hasta el surgimiento del 4G. Por último, la quinta generación comenzó a desarrollarse en torno a 2015, época en la que la ITU-R (*Unión Internacional de Telecomunicaciones*) publicó las primeras características del 5G. La ITU define tres tipos de servicios principales que proporciona el uso de 5G:

- ***Enhanced Mobile Broadband (eMBB)***: servicio que ofrece velocidades de transmisión de datos extremadamente altas (1Gbps para enlaces descendentes en interiores y 300 Mbps en exteriores) con una cobertura mejorada (permite hasta 1 millón de conexiones por kilómetro cuadrado) y con servicios de altas transferencias

cuando el usuario final presenta una gran movilidad (hasta 500km/h en trenes y hasta 1000km/h en aviones). Estos resultados se consiguen debido al aumento del ancho de banda que ofrece 5G. Por lo tanto, soporta una demanda cada vez más creciente de usuarios finales.

- **Massive Machine-Type Communications (mMTC):** servicio que proporciona conectividad con pocos requisitos de software y de hardware a los dispositivos que intervienen en la comunicación. Además, permite un funcionamiento de bajo consumo de energía y ahorro de batería. Se producen comunicaciones eficientes, fiables y seguras con un número elevado de usuarios finales (30.000 dispositivos por celda), donde los requisitos de latencia mínimos y con pequeños volúmenes de datos. Para permitir el ahorro de energía, típicamente, la movilidad de los dispositivos en la red es nula o muy reducida.
- **Ultra-reliable and low-latency communications (URLLC):** servicios que garantizan los requisitos muy críticos de determinadas aplicaciones en términos de latencia extremo a extremo (E2E), llegando incluso a valores de milisegundos, confiabilidad y disponibilidad de la red. Este tipo de comunicaciones se pueden expandir a multitud de nuevas aplicaciones. Para ello, se suelen transmitir paquetes de datos de tamaño reducido y de forma esporádica.

Seguidamente fue el 3GPP (*3rd Generation Partnership Project*) la organización encargada de estandarizar la parte radio y la arquitectura de un sistema 5G. El objetivo principal de esta organización es garantizar la interoperabilidad entre los elementos de la red independientemente del fabricante. Para ello, se necesita dar a conocer una especificación global de los estándares de comunicaciones móviles que se utilizan.

Con el transcurso del tiempo, 3GPP ha publicado diferentes versiones donde se estandarizan y se agregan mejoras de las funcionalidades de 5G. En 2018 se publica el 3GPP *Release 15* donde se indican los componentes básicos que forman la arquitectura de 5G y las características de las comunicaciones de banda ancha móvil (eMBB), incluyéndose mejoras en el segundo estándar de 5G 3GPP *Release 16* (2020). Pero este *Release* se centra principalmente en la definición de las comunicaciones de baja latencia ultra fiables (URLLC) y su integración con TSN para la automatización de la industria, además de las comunicaciones V2X (*Vehicle to everything*). Actualmente la organización está trabajando en la 3GPP *Release 17*, que se centra en la reducción de la complejidad de la parte radio, la realidad aumentada y una mejor precisión de la posición.



Figura 2: Evolución de las versiones publicadas por el 3GPP [2]

El 5G combina distintas tecnologías de radio, es decir, se implementa una nueva tecnología de interfaz de radio denominada *New Radio (NR)*, que se trata de la evolución de LTE.

Con el uso del 5G se pueden conseguir redes dedicadas y confiables para el uso en servicios locales, una conectividad inalámbrica escalable, y la capacidad de crear nuevos casos de usos. Pero para ello, es necesario la creación de redes privadas que no dependan de la red pública y estén formadas por su propio equipamiento con la utilización de múltiples tipos de espectros (con licencia, compartido y sin licencia) y con una comunicación de baja latencia ultra confiable (URLLC). Todas estas capacidades y funcionalidades están descritas en el 3GPP *Release 16*, ya que en el 3GPP *Release 15* se explica principalmente 5G *New Radio*.

### **1.2.2 Time-Sensitive Networking (TSN)**

*Time Sensitive Networking (TSN)* es un conjunto de estándares abiertos de la capa 2 que está especificado por el IEEE 802.1. Estos estándares resuelven los aspectos críticos de la Industria 4.0 con el objetivo de permitir el uso de Ethernet en aplicaciones sensibles al tiempo y admitir transmisiones deterministas con estrictos requisitos de latencia, confiabilidad y tiempo real. Por lo tanto, permite el transporte de mensajes de aplicaciones que presentan calidades de servicio (*Quality of Services, QoS*) diferentes, principalmente con unas variaciones de retardo muy bajos y con una escasa pérdida de datos. Es decir, esta tecnología es una red cableada basada en Ethernet y en el protocolo IEEE 802.1Q. Su principal objetivo es agrupar todas las tecnologías existentes de la capa 2. TSN emplea para la transmisión de paquetes la técnica de acceso al medio TDMA (*Time Division Multiple Access*) cuando se conoce el momento exacto de transmisión. Pero puede suceder que se desconozca entonces se utiliza multiplexación estadística (TDM asíncrona). La principal diferencia entre ambas técnicas se deba a que con TDMA se pre-asigna las ranuras temporales según el número de comunicaciones, independientemente si están activas o no. Mientras que con multiplexación estadística se detecta las comunicaciones activas y se asigna los *slots* entre las activas. De esta forma, la utilización del canal se incrementa al no desaprovechar los *slots* asignados a comunicaciones inactivas.

Una red TSN está formada principalmente por dispositivos finales, puentes e interfaces red-usuario. Los dispositivos finales son aquellos nodos fuente y destino de los flujos TSN, es decir, aquellos que ejecutan las aplicaciones y los servicios que requieren transmisiones deterministas. Los puentes son conmutadores Ethernet que transmiten las tramas Ethernet de un flujo TSN y las reciben siguiendo una programación de tiempo definido. Una interfaz de red-usuario (*User-Network Interface, UNI*) es la conexión entre el plano de usuario y el plano de control. En una UNI el lado de usuario está conformado por los emisores y receptores mientras que el lado de la red está formado por los puentes que transmiten las tramas de datos desde un emisor hasta uno o varios receptores. Por último, los flujos TSN son las tramas de tiempo crítico unidireccionales que se producen entre los dispositivos finales con unos requisitos de tiempo críticos determinados, presentando unos identificadores únicos para cada dispositivo final.

Las redes TSN pueden ser síncronas o asíncronas. En las redes TSN síncronas los relojes de todos los nodos de la red están sincronizados con la misma referencia de tiempo precisa mientras que en las asíncronas no necesitan la sincronización de sus dispositivos (puentes TSN). El objetivo de ambas redes es satisfacer los requisitos de QoS exigidos por las comunicaciones deterministas, pero en su implementación se cumplir o no. Sin embargo, las redes TSN asíncronas sufren un incremento de latencia, aunque mejoran la escalabilidad de la red y la utilización de los enlaces.

Estas mejoras que se producen en las redes TSN asíncronas se debe a no ser necesario la coordinación del tiempo en toda la red para programar la transmisión del tráfico TSN en instantes de tiempo reservados (*slots*).

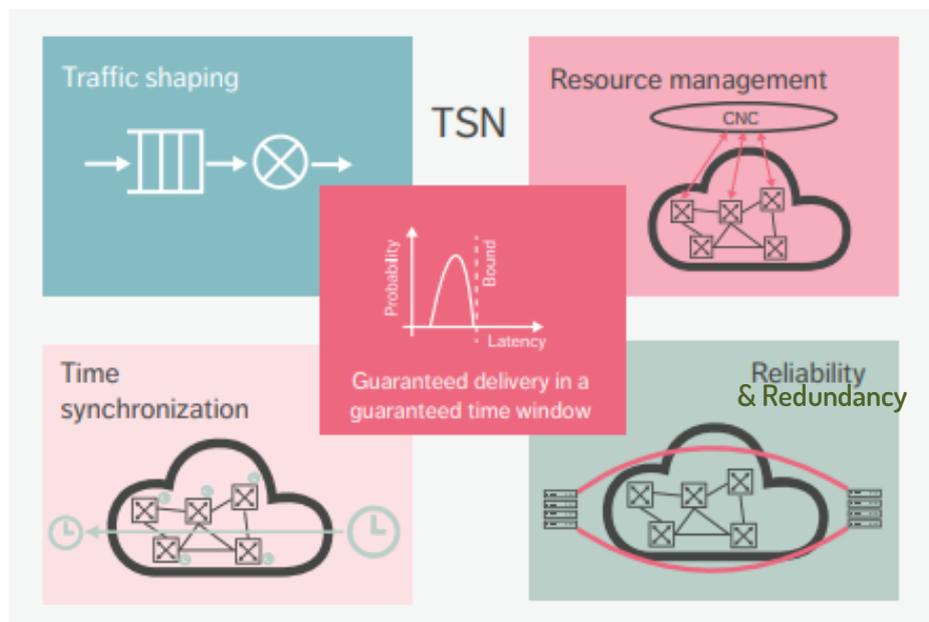


Figura 3: Características TSN [15]

En la Figura 3 se resumen los principales mecanismos que presentan las redes TSN, agrupadas en 4 bloques: conformación de tráfico, gestión de recursos, sincronización temporal y fiabilidad. Cada uno de estos bloques agrupa aquellos estándares IEEE TSN que permiten desarrollar dichas características. Además, estas son las herramientas que permiten el uso de TSN en la automatización industrial.

- Conformación de tráfico (*Traffic Shaping*):** en este bloque se agrupan aquellos estándares que garantizan la latencia y el *jitter* de una red TSN. Esto se consigue con la separación del tráfico en clases de tráfico según las calidades de servicio (QoS) requeridas. Es decir, la clase de tráfico (*Traffic Class, TC*) es la agrupación de los tráficos de los distintos tipos de servicios que contempla el *scheduler* dependiendo de sus características y requisitos, siendo tratados de forma diferentes según estos. Además, las tramas se reenvían y se clasifican en colas diferenciadas según estas clases de tráfico. En las redes TSN síncronas el *scheduler* de tráfico empleado está definido en el estándar IEEE 802.1Qbv. Mientras que en las redes TSN asíncronas el *scheduler* implementado se define en el IEEE 802.1Qcr (*Asynchronous Traffic Shaping*). Estos mecanismos determinan la configuración implementada en cada uno de los tráficos salientes de la interfaz de salida del switch/puente y filtran los flujos según su QoS, asignándoles unas prioridades. La principal diferencia es que el *scheduler* de las redes síncronas también determina el periodo de tiempo de transmisión de cada una de las tramas para calcular el retardo que se produce en dicho enlace de la red y poder configurar los relojes. Otro de los mecanismos empleados es *Ethernet frame preemption*, definidos en IEEE 802.3br y IEEE 802.1Qbu. Este mecanismo permite a las tramas de alta prioridad interrumpir la transmisión de las tramas de menor prioridad, reduciendo así el retardo de transmisión de las tramas críticas.

- **Resource Management:** la gestión de los recursos se realiza de manera diferente según el modelo de configuración TSN de la red. El 3GPP en el estándar IEEE 802.1Qcc describe tres modelos de red TSN, explicados en la Sección 2.2 TSN. Sin embargo, independientemente de la arquitectura del plano de control de la red TSN, la gestión de recursos permite el descubrimiento dinámico de los nodos que conforman la red TSN, así como la topología implementada. A su vez proporciona la configuración y la monitorización de la red. También se encarga de la asignación y el registro de los recursos que son necesarios para garantizar los requisitos de cada flujo de tráfico. Algunas de estas propiedades se implementan con los métodos y técnicas definidas en los siguientes estándares del IEEE: la reserva de los flujos de trama se define en el IEEE 802.1Qat y en el IEEE 802.1Qcc mientras que la reserva del enlace local en el IEEE 802.1CS.
- **Reliability & Redundancy:** la alta fiabilidad y disponibilidad de estas redes se consigue con la utilización del mecanismo *Frame Replication and Elimination for Reliability (FRER)*, definido en el estándar IEEE 802.1CB, para los flujos de datos con confiabilidad a nivel de paquete. Es decir, se transmiten múltiples copias de una misma trama, pero por caminos disjuntos de la red, evitando la interrupción de la comunicación por la caída de algún enlace, por ejemplo. La gestión de los caminos es una característica esencial para evitar problemas en la reserva de los mismos, por ejemplo, para saber si están disponibles. Para ello se implementan la técnica de control y la reserva de los caminos (*Path Control & Reservation*) definido en el estándar IEEE 802.1Qca. Además, se implementan técnicas de filtrado y políticas de control para cada flujo (802.1Qci), reduciendo el retardo y evitando un mal funcionamiento de la red o la violación del ancho de banda.
- **Time Synchronization:** Cada uno de los componentes de la red TSN presenta un reloj propio a nivel de capa 2 del modelo OSI. Estos relojes inicialmente tienen la misma referencia de tiempo, pero pueden sufrir desviaciones que se acumulan y provocan el mal funcionamiento de la red. Para evitar este problema y sincronizar todos los relojes de la red se utiliza el mecanismo *generalized Precision Time Protocol (gPTP)*, que es un perfil del estándar *Precision Time Protocol (PTP)* definido en IEEE 802.1AS. Este mecanismo entrega el tiempo apropiado de un dispositivo a otro teniendo en cuenta las posibles desviaciones que se pueden producir. Para ello, primeramente, se estima la latencia que se introduce en la red con el intercambio de una serie de mensajes y seguidamente, se usa esta latencia para sincronizar los relojes de los distintos nodos. De esta forma, todos los nodos de la red presentan la misma noción de tiempo para la asignación de los intervalos de tiempo a cada flujo (transmisiones TDMA), garantizando una llegada precisa entre ambos extremos de la comunicación.

### 1.3 Motivación

La constante evolución que se está produciendo en los diferentes sectores principalmente en el industrial para conseguir comunicaciones deterministas de tiempo real es el principal motivo para el estudio que se está llevando a cabo. Cada vez son más las empresas tecnológicas que apuestan por la integración de la tecnología TSN con 5G para permitir la automatización industrial, siendo actualmente la primordial vía de estudio de alianzas, empresas e instituciones como 3GPP (*3rd Generation Partnership Project*), 5G-ACIA (*5G Alliance for Connected*

*Industries and Automation*), Ericsson, Cisco, Qualcomm, etc. Así como grupos de investigación para solucionar los problemas que surgen de su integración.

Por otro lado, la mayoría de las investigaciones que se están llevando a cabo emplean TSN síncrono como la mejor solución para la Industria 4.0. Sin embargo, tienen una escalabilidad limitada debido a los mecanismos de sincronización necesarios y son complejas de configurar y gestionar. Asimismo, el proceso de sincronización y monitorización de los relojes de la red consume recursos de ancho de banda en la red y la estrategia TDMA para orquestar las transmisiones de los distintos flujos puede resultar en una utilización pobre de los recursos de la red cuando hay dominancia de tráfico aperiódico. Por tanto, puede ser una apropiada solución optar por TSN asíncrono (no necesita una referencia de tiempo común entre todos los nodos de la red industrial) para ciertos escenarios industriales como aquellos que presentan dominancia de tráfico aperiódico determinista como, por ejemplo, aquel que generan los vehículos guiados automáticamente. El principal problema de TSN asíncrono es que no es capaz de conseguir retardos tan reducidos como TSN síncrono, pero según los estudios llevados a cabo con esta tecnología en [17] y [18], los retardos experimentados para los distintos tipos de tráfico se ajustan a los retardos requeridos para la mayoría de los servicios que se esperan en la Industria 4.0.

Además, según se ha especificado en el 3GPP *Release 16* la integración entre 5G y TSN síncrono se ha considerado teniendo inicialmente una red TSN desplegada en la industria incorporando, posteriormente, la tecnología 5G. Por lo que, los nuevos nodos agregados con la tecnología 5G actuaría formando un puente TSN virtual que conectaría dos redes TSN o una red TSN con los dispositivos finales, donde todos los mensajes TSN se transmitirían encapsulados por la red 5G. Sin embargo, existen muy pocas investigaciones, solamente se ha estudiado en [4] sobre cómo se puede habilitar la funcionalidad de *5G Network Slicing* a través de una red TSN. Esta característica permitirá aislar los diferentes tipos de tráfico que se transmiten por la red TSN-5G en *slices* con unos requisitos de tráfico determinados y sin interferencias de rendimiento entre ellas.

Por tanto, al ser escaso el número de trabajos realizados sobre *network slicing* en redes TSN y considerando que es un tema de especial relevancia para aumentar la utilización de la red física, este trabajo se centra en la implementación de la red de transporte TSN asíncrono para redes industriales privadas 5G. Es decir, se busca definir una solución para el soporte de *network slicing* en las redes de transporte 5G industriales. Se opta por la utilización de TSN asíncrono en lugar de TSN síncrono para implementar esta red de transporte. Además, una vez resuelto el problema de soporte se estudia la granularidad, en términos de número de *slices* adecuadas, para dar servicio al tráfico de la Industria 4.0 y la validación del grado de aislamiento requerido por los servicios industriales en la red de transporte.

### **1.4 Objetivos**

El objetivo de este trabajo, como se ha comentado anteriormente, es la implementación de una red de transporte TSN asíncrona sobre una red 5G con la finalidad de ofrecer conectividad de baja latencia y de alta fiabilidad entre los distintos dispositivos finales que conforman las redes de la Industria 4.0. Además, estas redes permitirán el proceso de automatización, monitorización y control en tiempo real de los diferentes nodos de la red. El anterior objetivo se puede descomponer en los siguientes subobjetivos específicos:

- **Estudiar las características de las tecnologías 5G y TSN.** Es necesario conocer el funcionamiento y los mecanismos que se implementan tanto en la tecnología de quinta generación de comunicaciones *wireless* como en *Time-Sensitive Network* para comprender la integración que se desea realizar.
- **Identificar los requisitos tanto funcionales como de rendimiento para la arquitectura que integre las tecnologías 5G y TSN.** Una vez conocido el funcionamiento de la tecnología 5G y TSN, se identifica los principales servicios que ofrece la Industria 4.0 para definir las funcionalidades requeridas y las restricciones en el rendimiento de la red. Además del estudio de los diseños de arquitectura de integración de 5G y TSN desarrollados en la literatura.
- **Estudio de la funcionalidad de *network slicing* en redes TSN asíncronas y su integración como red de transporte en una red 5G.** Una vez cumplido el anterior objetivo, se estudia la mejor solución para compatibilizar *network slicing* que ofrece las redes 5G en las redes TSN asíncronas. Además, se definen las principales propiedades de los tipos de tráfico que intervienen en la Industria 4.0 para su posterior estudio en distintas *slices*.
- **Estudio de la granularidad de las *slices* que conforman una red de transporte TSN asíncrona.** A través del simulador de una red TSN asíncrona se estudia la granularidad que pueden presentar las *slices* y cómo afectan a los retardos requeridos extremo a extremo para cada uno de los tipos de tráfico, así como a la probabilidad de rechazo de cada *slice*.
- **Desarrollo de la documentación y de la memoria.** Por último, se redacta detalladamente toda la información recopilada y los resultados obtenidos en cada una de las simulaciones para el desarrollo del trabajo.

### 1.5 Organización

En la Sección 1.4 *Objetivos* se ha detallado cada uno de los objetivos secundarios establecidos para conseguir alcanzar el objetivo final de la integración entre 5G y TSN asíncrono. Cada uno de los anteriores objetivos detallados se han alcanzado mediante la realización de una o varias tareas que se especifican y describen en este apartado.

- **Tarea 1. Estudio del funcionamiento de la tecnología inalámbrica de quinta generación.** El primer paso es adquirir los conocimientos necesarios sobre el funcionamiento y la arquitectura de las redes 5G, así como sus principales características. Para ello, se revisa a la literatura existente hasta la fecha sobre 5G.
- **Tarea 2. Estudio del funcionamiento de la tecnología *Time-Sensitive Network* (TSN).** También es necesario aprender el funcionamiento y la arquitectura de las redes TSN, recurriendo a la literatura actualmente existente.
- **Tarea 3. Estudio de las arquitecturas existentes de la integración entre TSN y 5G.** Se estudia los diferentes diseños propuestos en los distintos estudios e investigaciones para llevar a cabo la integración entre la tecnología TSN y 5G. Con ello, se desea deducir los principales requisitos y características que se deben tener en cuenta de

ambas tecnologías.

- **Tarea 4. Análisis de realización de *network slicing* en una red de transporte 5G basada en TSN.** Son reducidas las investigaciones que tratan esta cohesión entre las *slices* de los dominios de una red 5G y una red de transporte basada en la tecnología TSN. Además, no se han estudiado las implicaciones de considerar una red TSN asíncrona en este contexto.
- **Tarea 5. Familiarización con el software de simulación de una red TSN asíncrona.** A partir del simulador de una red TSN asíncrona proporcionado, se ha estudiado las diferentes posibilidades que ofrece. Este simulador contempla diferentes tipos de topologías, se pueden definir una gran variedad de tipos de tráfico con distintos requisitos de *QoS*, así como se puede resolver la red a través de 4 soluciones posibles del optimizador (Neptuno, Baseline, Mindelay u Optimal). En este caso se opta por la solución óptima que proporciona los mejores resultados para configurar la red TSN asíncrona de modo que se minimice la probabilidad de rechazo de flujo.
- **Tarea 6. Definir posibles escenarios de estudio de la granularidad.** Esta tarea se refiere a definir de forma precisa los experimentos y configuraciones necesarias para dar respuesta a las cuestiones abordadas en este trabajo. El simulador permite definir *slices* con distintas propiedades de tráfico en una red TSN, por lo que se definen determinados casos de uso para comprobar el funcionamiento. Así como se estudia la granularidad necesaria en cuanto a número de *slices* necesarios en la red de transporte basada en TSN para poder acomodar tráfico con una gran variedad de requisitos de *QoS*.
- **Tarea 7. Análisis de los resultados obtenidos con los experimentos de granularidad.** Los escenarios definidos en la anterior tarea se simulan en el software y se comprueba la cantidad de tráfico con distintos requisitos de *QoS* que puede acomodar una *slices* dada y si se cumple con las *QoS* demandas.
- **Tarea 8. Documentación.** Para elaborar la memoria de este trabajo se ha utilizado el editor de textos Microsoft Word, ya que facilita un formato y un diseño apropiado para este tipo de trabajo.

Por último, la mayor parte las tareas se han realizado de manera secuencial según el orden en el que se han descrito, aunque algunas de ellas se han desarrollado de forma paralela por ser ortogonales entre sí. A continuación, se añade un diagrama de Gantt donde se visualiza la progresión temporal de manera detallada del proyecto.

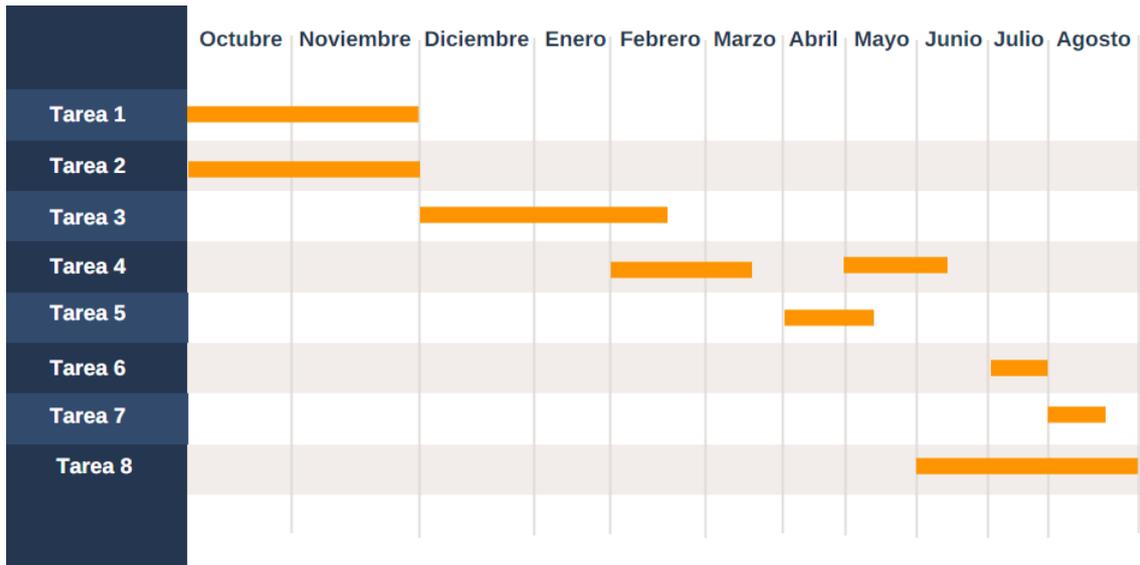


Figura 4: Temporización del proyecto.



# Capítulo 2

## Estado del arte

En este apartado se detallan las principales características y el funcionamiento de las tecnologías de 5G y TSN. Así como la arquitectura de cada una de ellas y los diferentes diseños que se han desarrollado de su integración, especificados principalmente en el 3GPP *Release 16*.

### 2.1 5G

Con la llegada del 5G la velocidad de datos se ha incrementado 100 veces más que la disponible con 4G, y la cantidad de dispositivos conectados [32]. Pero no solamente afecta a las anteriores características, sino que también aumenta la disponibilidad de la red al 100% y la cobertura geográfica de manera segura y con una privacidad mejorada. Además, se reduce el consumo de energía de las baterías que utilizan los diferentes dispositivos y nodos 5G, extendiendo su vida útil [2].

Aparte de las anteriores características, una de las principales finalidades de 5G es la capacidad de construir industrias verticales. 5G actuará como la red de comunicaciones inalámbricas que es necesaria en la infraestructura de IoT para las industrias verticales. En estas industrias se producirán nuevos casos de uso con distintos requisitos de QoS que debe admitir 5G. Esto se logra gracias a las comunicaciones ultra confiables de baja latencia (URLLC) que implementa 5G, ya que permite la transmisión de paquetes de pequeño tamaño dentro de un límite de tiempo determinado. Al proporcionar la amplia gama de casos de uso los sistemas 5G operan en un rango de frecuencia muy amplio desde menos de 1 GHz hasta frecuencias de ondas milimétricas (mmW). Estas bajas frecuencias se utilizan para ofrecer latencias muy bajas, alta confiabilidad y altas velocidades de datos en áreas amplia, mientras que las implementaciones con frecuencias superiores a 10GHz se utilizan para ofrecer servicios con velocidades de datos y capacidades extremas en áreas densas. También emplean múltiples antenas direccionales de formación de haces con un esquema MIMO (*Multiple Input Multiple Output*) y se produce la división entre el plano de control y el usuario.

El sistema 5G se divide principalmente en dos componentes de red: la red de acceso por radio (RAN, *Radio Access Network*) y la red central (CN, *Central Network*). La parte RAN está compuesta por el equipo usuario (*User Equipment*, UE) y el gNB (*new generation Node B*). El 5G CN está formado por funciones de red independientes para el plano de usuario (*User Plane*,

UP) y el plano de control (*Control Plane*, CP). Por un lado, el UP se encarga de transportar el tráfico de usuario de la red mientras que el CP transporta el tráfico de señalización y control. La función que maneja las rutas del UP entre el gNB y la red de datos externas, que puede ser una red TSN, es la función del plano de usuario (*User Plane Function*, UPF). Mientras que en el plano de control existen diferentes funciones, como la función de gestión de sesión (*Session Management Function*, SMF) que interactúa con el UP para establecer o modificar las sesiones de PDU (unidad de datos de protocolo). Estas sesiones PDU son las encargadas de definir los servicios de conectividad entre UE y la red de datos. Otra función es la función de control de políticas (*Policy Control Function*, PCF) que proporciona las reglas de políticas a la SMF. Por última, está la función de aplicación (*Application Function*, AF) que interactúa con la PCF para el control de las políticas e influye en el enrutamiento del tráfico en el UP.

La baja latencia de la red se consigue reduciendo los tiempos de procesamiento en UE y gNB a través de la utilización de tramas más flexibles y ranuras de transmisión más cortas con distintos espaciados entre subportadora. Por tanto, los slots presentan una duración de 1 ms a 0.0625 ms, siendo posible transmitir la información usando solo una parte de la ranura, denominada *minislot*. La latencia también se reduce con el mecanismo de programación sin concesión para las transmisiones de enlace ascendente. Es decir, el gNB pre-asigna unos recursos al UE y cuando el UE desea transmitir, este no necesita realizar la reserva de recursos al gNB, reduciendo el retardo asociado a este proceso, y transmitiendo los datos inmediatamente. Sin embargo, como se ha comprobado solo la parte radio ya presenta un retardo de 1 ms que se debe incrementar debido a los retardos que se originan en la parte de transporte, siendo por tanto imposible cumplir con los retardos demandados extremo a extremo con el uso de esta tecnología solamente.

Por tanto, la reducción del retardo para permitir una comunicación más interactiva se realiza tanto en la parte radio de 5G como en las tecnologías que interfieren en las comunicaciones extremo a extremo. Para lograr los requisitos de latencia y retardo la arquitectura de la red 5G ha sufrido una evolución, como se muestra en la Figura 5. Originalmente las estaciones base 5G (gNB) se encargaban tanto de la parte radio como del procesamiento en banda base de las transmisiones. Pero con la llegada de la “nube” se produjo el desacoplo de estas funcionalidades en los gNBs. Las funciones de la capa física, como la frecuencia radio (RF), la amplificación de la señal o los procesos de conversión D/A y A/D, son realizadas en el cabezal de radios remota (*Head Radio Remote*, RRH). Mientras que los protocolos de las capas superiores y el procesamiento en banda base se realizan en un conjunto centralizado de unidades de banda base (*Base Band Units*, BBUs), denominando al enlace de unión entre el RRH y el BBU *fronthaul*. Sin embargo, es necesario en el enlace *fronthaul* un ancho de banda extremo para desarrollar las funciones de la BBUs. Por consiguiente, gracias a la tecnología SDN (*Software Defined Network*) y NFV (*Network Function Virtualization*) junto con la división funcional se virtualizó la red de acceso radio (RAN) en la nube y se utilizan nuevas tecnologías como *Multi-access Edge Computing* (MEC). La tecnología MEC permite trasladar toda la computación y las aplicaciones a las centrales locales de los diferentes operadores, eliminándose los posibles retardos que se pueden introducir en las redes WAN. Es decir, las funciones RAN, como la modulación y la demodulación, se trasladan del grupo de BBU centralizado a las RRHs. El cabezal de radio remoto (RRH) evoluciona a una unidad de antena activa (*Active Antena Unit*, AAU). Mientras que la BBU se divide en una unidad centralizada (*Centralized Unit*, CU) virtualizada y una unidad distribuida (*Distributed Unit*, DU) para implementar las funciones RAN en distintas ubicaciones de manera flexible. De este modo, las funciones de la capa física son desarrolladas por la unidad radio (*Radio Unit*, RU), la DU realiza los protocolos de las capas localizadas entre la RU y la CU,

y la CU implementa los protocolos de las capas superiores. La CU puede estar conectado a varias DU.

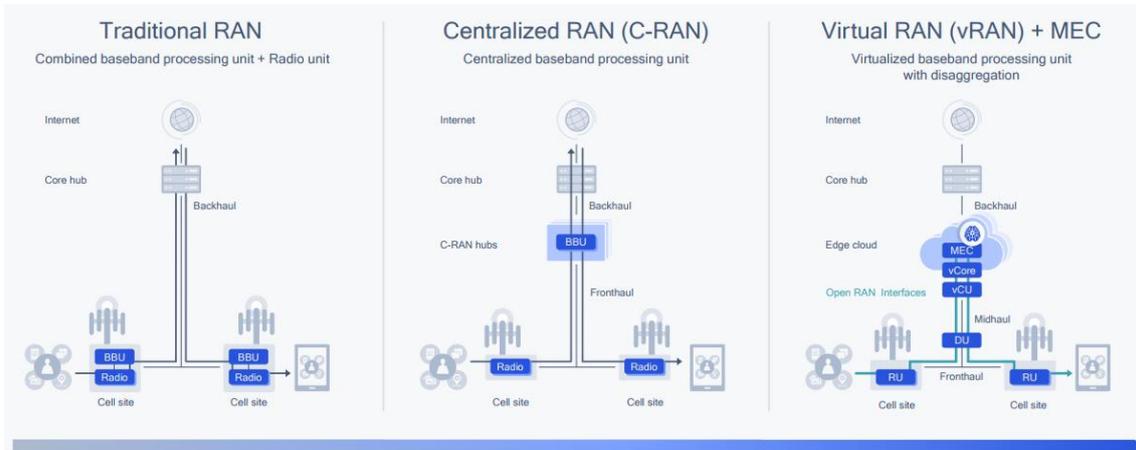


Figura 5: Evolución de la red 5G. [19]

De esta forma, los datos son procesados cerca de sus puntos de origen, exactamente en los gNBs, reduciendo considerablemente el retardo. También soporta múltiples servicios con distintas necesidades de latencia, *throughput* y fiabilidad, e incrementa la seguridad y la privacidad de los datos al almacenarlos localmente. Asimismo, se aumenta la disponibilidad y la escalabilidad de la red a través de escalar de forma independiente los recursos disponibles para el tráfico del plano de control y el del usuario.

En las redes 5G se introduce, como se ha comentado anteriormente, las redes definidas por software (*Software Defined Networking, SDN*) y la virtualización de la función de red (*Network Function Virtualization, NFV*). Las redes definidas por software permiten programar redes con unas determinadas características personalizadas y optimizadas para diferentes servicios dentro de una misma arquitectura física. Por tanto, con SDN se permite abstraer de la infraestructura de red física y cambiar el comportamiento de la red según las capacidades que desee un determinado tráfico gracias a la programación. De tal manera que con SDN se simplifica la gestión de las redes. Por otro lado, NFV permite implementar una función de red a través de la programación en lugar de utilizar una pieza física hardware gracias a la separación que se ha producido entre el hardware y el software. De esta forma, se aumenta la escalabilidad de la red y la flexibilidad de ejecutar una función de red independientemente de su ubicación. Esto es posible ya que una vez que se virtualiza una función de red esta no está vinculada a ningún nodo específico o a una ubicación, permitiendo ejecutar una misma función de red en diferentes lugares para distintos segmentos de red.

Por otro lado, como se ha comentado en la Sección 1.2.1 *Redes de banda ancha de Quinta Generación (5G)*, la ITU define 3 tipos principales de servicios 5G (eMBB, mMTC e URLLC). Estos servicios presentan diferentes requisitos de calidad de servicio entre ellos y a su vez, cada uno puede albergar una gran variedad de casos de usos. En 5G se ha desarrollado una funcionalidad, denominada *network slicing* que permite dividir la arquitectura física de la red en bloques de construcción lógicos (*slices*) con los requisitos necesarios para adaptarse a los servicios que ofrecen y escalar. Esto se consigue con el uso de la tecnología de nube junto con SDN y NFV. De este modo, la red puede ir formando *slices* adaptándose a las demandas cambiantes de los suscriptores y al cumplimiento de los requisitos de los nuevos casos de uso. Estas características se explican más en detalle en la Sección 3.1 *Network Slicing en 5GS*.

Los sistemas 5G admiten flujos con diferentes QoS según las exigencias del usuario. Un flujo con QoS en los sistemas 5G se caracteriza dependiendo del tipo de recursos necesario para su implementación. Por tanto, el flujo QoS puede tener una tasa de bit garantizada (*Guaranteed Bit Rate*, GBR) o no. Si el flujo es GBR entonces se pueden emplear recursos para garantizar un retardo crítico, es decir, el flujo es de tipo *Delay Critical* GBR. Cada flujo QoS está relacionado con un perfil QoS 5G que contiene:

- Prioridad de asignación y retención (*Allocation and Retention Priority*, ARP).
- Tasa de bit de flujo garantizada (*Guaranteed Flow Bit Rate*, GFBR).
- Tasa de bit de flujo máxima (*Maximum Flow Bit Rate*, MFBR).
- Tasa de pérdida de paquetes máxima (*Maximum Packet Loss Rate*).
- Identificador QoS 5G (*5G QoS Identifier*, 5QI).

El 5QI referencia a un conjunto de características de QoS, que son:

- **Nivel de prioridad:** el nivel de prioridad determina la prioridad del flujo en el *scheduler* para acceder al medio compartido. La prioridad más alta se atribuye al valor más bajo, donde se asigna un nivel diferente a los flujos con distintos requisitos.
- **Presupuesto de retardo de paquete** (*Packet Delay Budget*, PDB): describe la latencia máxima de un paquete dentro del sistema durante su transmisión y se divide según los componentes de red del sistema 5G, es decir, RAN PDB para la conexión entre el UE y el UPF, y CN PDB para la conexión en el plano de control.
- **Tasa de error de paquete** (*Packet Error Rate*, PER): define la cantidad máxima de paquetes que un nodo 5G puede enviar y procesar pero que no llegan a su destino.
- **Volumen de ráfaga de datos máxima** (*Maximum Data Burst Volume*, MDBV): define la cantidad máxima de datos que el RAN debe servir dentro de un periodo de RAN PDB.
- **Ventana promedio** (*Averaging Window*): indica el tiempo necesario para calcular el GFBR y el MFBR.

Las características que componen el 5QI pueden estar preconfiguradas en el RAN o se pueden definir dinámicamente. En [9] se establece una tabla con el mapeo de 5QI a las características de QoS 5G según los tipos de servicio. Estos valores de 5QI pueden ser asignados estáticamente en el RAN. En la Tabla 1 se muestra los valores de QoS para los 5QI, donde la mayoría de las entradas son de retardo crítico GBR, porque son los principales tipos de tráfico empleados en la Industria 4.0 [10].

5QI Value	Resource Type	Priority Level	PDB	PER	MDBV	Averaging Window	Example Services
80	Non-GBR	68	10 ms	$10^{-6}$	N/A	N/A	Low Latency eMBB applications Augmented Reality
82	Delay-critical GBR	19	10 ms	$10^{-4}$	255 bytes	2000 ms	Discrete Automation
83		22	10 ms	$10^{-4}$	1354 bytes	2000 ms	Discrete Automation, V2X messages-
84		24	30 ms	$10^{-5}$	1354 bytes	2000 ms	Intelligent transport systems
85		25	5 ms	$10^{-3}$	500 bytes	2000 ms	Interactive Service - Motion tracking data
86		25	10 ms	$10^{-3}$	1125 bytes	2000 ms	Interactive Service – Motion tracking data

Tabla 1: Mapeo de los 5QI a las características de 5G QoS

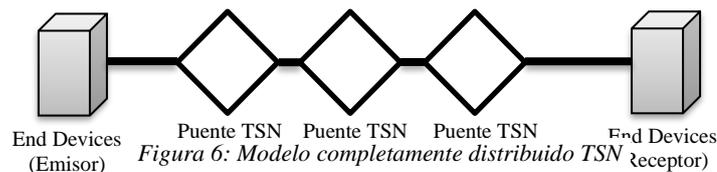
## 2.2 TSN

La tecnología *Time-Sensitive Network* (TSN) presenta mecanismos de *scheduler* distintos dependiendo de si los relojes de los nodos tienen la misma referencia de tiempo o no, como se comentó en la Sección 1.2.2 *Time-Sensitive Networking* (TSN). Es decir, si los nodos presentan la misma referencia temporal se utiliza el modelo de TSN síncrono mientras que si no se realiza una sincronización temporal de los nodos entonces se utiliza el modelo de TSN asíncrono. En este apartado se desarrollan las principales características que comparten ambos modelos, y posteriormente las características que diferencian al modelo TSN síncrono y TSN asíncrono. Por último, se detallan las ventajas e inconvenientes de TSN síncrono y asíncrono.

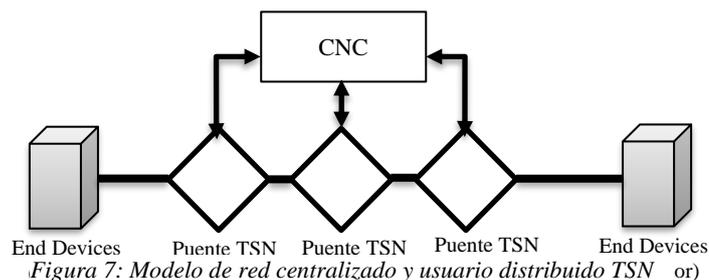
TSN presenta tres arquitecturas de configuración en el plano de control para la asignación, la configuración, el registro y la gestión de los recursos. Uno de los objetivos del uso de TSN es que se determine los requisitos de los flujos sin necesidad de tener conocimiento de la propia red. Estos requisitos son obtenidos por la red y analizan la topología de la red y las capacidades de los puentes para posteriormente configurar los puentes para cumplir con los requisitos determinados. La forma de obtención de los requisitos del tráfico depende de los modelos de configuración de la red TSN. Estos modelos de configuración son: completamente centralizado, completamente

distribuido, y un modelo de configuración de red centralizado y usuario distribuido; y se definen en el estándar IEEE 802.1Qcc, centrado principalmente en el modelo completamente centralizado.

- **Modelo completamente distribuido:** Los requisitos de los usuarios de las estaciones finales se propagan por toda la topología a través del uso de un protocolo distribuido. Es decir, cada componente de la red comparte las propiedades requeridas al resto de nodos de la red para establecer los flujos TSN con QoS. En este modelo la interfaz de red-usuario se localiza entre las estaciones finales y el primer o final puente TSN de la red. La Figura 6 muestra una representación visual de un modelo completamente distribuido de una red TSN.



- **Modelo de red centralizado y usuario distribuido:** En este modelo se introduce un componente denominado controlador de red centralizado (*Centralized Network Configuration, CNC*). El CNC tiene una visión completa de toda la topología de la red y de todas las tramas que se transmiten en la red. Esto es necesario ya que se encarga de la configuración de los puentes TSN y de operaciones complejas de rendimiento requeridas por el *Time-Aware Shaper (TAS)*, *Frame Preemption*, etc., en los puentes para la programación de las tramas y las rutas que deben seguir. La UNI se ubica aún entre el puente y la estación final. Sin embargo, en este modelo los requisitos de usuario se comunican desde el emisor hasta el puente de borde de la red (conectado a una estación final), que se lo comunica al CNC. En la Figura 7 se muestra la representación de un modelo de red centralizado y usuario distribuido de una red TSN.



- **Modelo completamente centralizado:** En este modelo además de utilizar la entidad CNC también considera la entidad controladora de usuario centralizado (*Centralized User Configuration, CUC*). El CUC se encarga de descubrir las estaciones finales y recibir sus características y los requisitos de los flujos TSN. La diferencia con el modelo anterior es que en este caso el intercambio de requisitos de usuario se realiza entre el CNC y el CUC, es decir, la UNI se ubica entre el CNC y el CUC. En la Figura 8 se visualiza un modelo completamente centralizado, donde los flujos de información son el siguiente:

1. Las entidades finales envían sus requisitos de QoS al CUC. Algunos de estos requisitos son la velocidad de los datos, las clases de tráfico, las prioridades y la latencia.
2. El CUC reenvía esta información al CNC a través de la interfaz de información de configuración red/usuario.
3. Los puentes TSN transmiten sus capacidades al CNC. Algunas de estas capacidades son los retardos de los puentes según el puerto y la clase, los retardos de propagación según el puerto y las prioridades que considera.
4. El CNC determina la configuración de cada puente con el objetivo de cumplir con los requisitos de los flujos TSN gracias a la información transmitida del CUC y de los puentes. Así como el cronograma de transmisión del tráfico según los tiempos de inicio de los flujos y los tiempos de control de apertura y cierre de puerta en los puertos TSN.
5. Una vez obtenida la configuración, el CNC puede enviar información de configuración TSN específica al CUC que, a su vez, este puede reenviar a las estaciones finales.

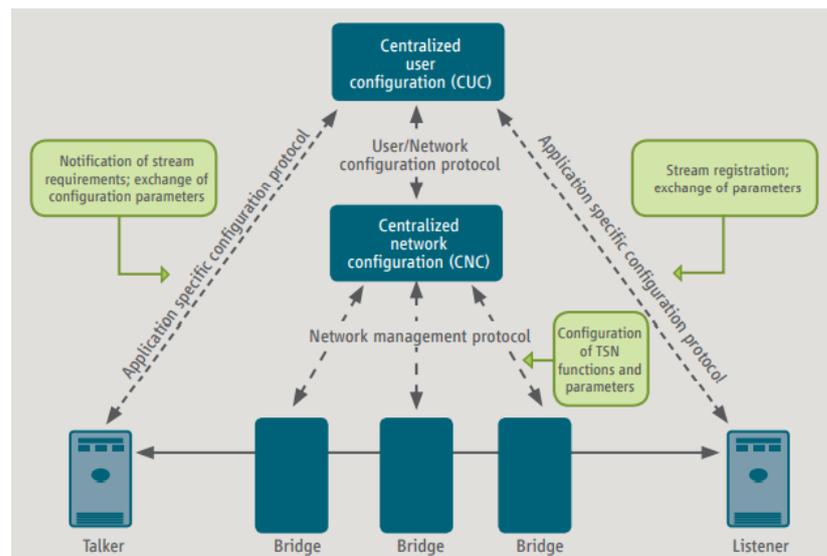


Figura 8: Modelo completamente centralizado TSN [29]

En estas redes puede coexistir múltiples CUC, pero solo una entidad CNC. Con la existencia del CNC y el CUC se implementa la automatización de la gestión de la red a través del paradigma de redes definidas por software (*Software Defined Networking, SDN*). De esta forma, se puede conocer si la red cumple con los requisitos de QoS previamente al establecimiento de la configuración de la red.

El modelo de arquitectura completamente distribuido, y la red centralizada y el usuario distribuido se basan en el protocolo de reserva de transmisión (*SRP, Stream Reservation Protocol*) que a su vez utiliza el protocolo de registro de transmisión múltiple (*MSRP, Multiple Stream Registration Protocol*), el protocolo de registro de VLAN múltiple (*MVRP, Multiple VLAN Registration Protocol*) y el protocolo de registro de MAC múltiple (*MMRP, Multiple MAC Registration Protocol*). Sin embargo, esta serie de protocolos no se pueden utilizar ya que el

protocolo SRP no cumple con las necesidades que se solicitan para las redes de automatización industrial. Se está estudiando y desarrollando nuevos protocolos para la implementación de estos modelos. Por consiguiente, este trabajo se centra en el modelo de TSN totalmente centralizado. Además, el modelo totalmente centralizado presenta ventajas frente a los modelos distribuidos, ya que los sistemas centralizados basados en SDN para la red de transporte se pueden incorporar fácilmente en los sistemas de gestión y control 3GPP existente.

El CNC necesita conocer la topología de la red completa para su funcionamiento, es decir, para la planificación y la optimización de las rutas de comunicaciones extremo a extremo. El protocolo que se utiliza para descubrir los enlaces entre los puentes y las estaciones finales de una red es el protocolo de descubrimiento de capa de enlace (*Link Layer Discovery Protocol*, LLDP). Este protocolo está estandarizado en el IEEE 802.1AB y permite tanto el descubrimiento de la topología como el estado de los dispositivos y su disponibilidad. En cada uno de las estaciones finales o puentes existe un conjunto de agentes LLDP que se encarga del intercambio de información entre los nodos vecinos. Estos agentes LLDP pueden recibir y procesar tramas LLDP, así como transmitirlos. Las tramas LLDP son específicas del puerto de salida para poder determinar la topología de la red. Estas tramas LLDP son recibidas por una estación final o un puente con un agente que se encargará de verificarlas y almacenar la información contenida en una base de información de mensajes (*Message Information Base*, MIB) remota. Toda la información almacenada puede ser recuperada por cada puente si se utiliza las interfaces de administración como las basadas en el protocolo de administración de redes simples (SNMP) o MIB. El intercambio de mensajes LLDP se representa en la Figura 9.

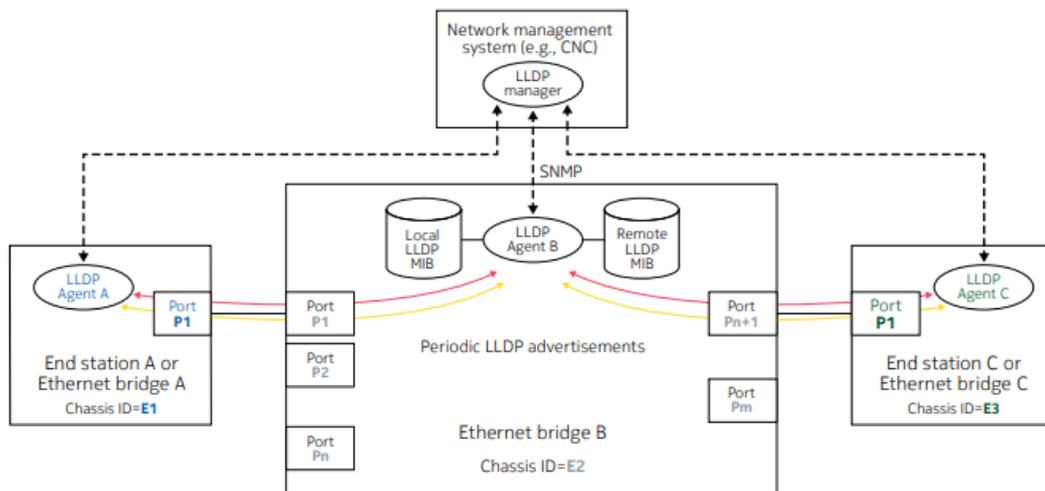


Figura 9: Intercambio de información LLDP [5]

El protocolo de administración de redes simples (*Simple Network Management Protocol*, SNMP) está basado en IP para la monitorización y la administración de los diferentes puentes de red y estaciones finales. En una red TSN centralizada se define una interfaz de gestión entre los puentes y el CNC para recibir la información sobre el estado y la configuración actual de los nodos de la red. Esta información es transmitida por cada uno de los agentes SNMP existente en los puentes o estaciones finales. Además, estos agentes SNMP pueden manipular el estado de los nodos de la red, así como dispara notificaciones al CNC, que es el sistema de gestión de la red, para informar sobre determinados eventos. En todos los procesos se emplea el uso de una MIB.

TSN divide el tráfico en ocho tipos de tráfico diferentes que se identifica con el campo de punto de código de prioridad (*Priority Code Point*, PCP) en la etiqueta VLAN de la cabecera

Ethernet. El PCP determina la prioridad de la trama, donde el “0” es tráfico de menor prioridad, siendo el tráfico *Best Effort (BE)* y la clase por defecto; y el “7” es el tráfico de mayor prioridad, como se observa en la Tabla 2 según el tipo de tráfico considerado en [29].

Tipos de tráfico	Valor de PCP	Criticidad
Control de red	7	Alta
Isócrona	6	Alta
Síncrono cíclico	5	Alta
Asíncrono cíclico	5	Alta
Eventos: control	4	Alta
Eventos: alarmas y comandos del operador	3	Media
Configuración y diagnóstico	2	Media
Vídeo	1	Baja
Audio y voz	1	Baja
Mejor esfuerzo	0	Baja

Tabla 2: Valores PCP según la criticidad y el tipo de tráfico [29]

El valor del PCP que se asigna a cada tráfico depende de los requisitos de QoS y criticidad. La criticidad es la cuantía de riesgo que presentan los datos que intervienen en el funcionamiento de las partes críticas de un sistema. Dependiendo de su nivel de gravedad se utilizan unos mecanismos de QoS/TSN apropiados, así como la cantidad de ancho de banda reservado y los instantes temporales en los que se transmite dicho tráfico. Existen tres niveles de criticidad, como se observa en la Tabla 2 para los tipos de servicios considerados en [29]:

- **Alto:** se emplea para tráfico de aplicaciones y servicios de red que son muy críticos para el funcionamiento del sistema. Si se produce la pérdida de alguno de los datos de estos tipos de tráfico se produce un mal funcionamiento del sistema y la aplicación no podrá repetir ni retransmitir de nuevo los datos perdidos. Este tipo de tráfico presenta un valor de PCP de 4 a 7.

- **Medio:** se emplea para tráfico de aplicaciones y servicios de red que son relevantes, pero no necesitan el funcionamiento constante de las partes críticas del sistema. Además, la pérdida de datos puede causar una degradación del funcionamiento, pero no el mal funcionamiento de todo el sistema y la aplicación si puede volver a retransmitir los datos perdidos. El valor de PCP asignado a este tipo de tráfico es 2 o 3.
- **Bajo:** se emplea para tráfico de aplicaciones y servicios de red que no son importantes para el funcionamiento de las partes críticas del sistema. Por lo tanto, la pérdida de estos datos no supone ningún problema para el correcto funcionamiento de las partes críticas del sistema y la aplicación puede retransmitir los datos. En consecuencia, se le asigna un PCP 0 o 1 a este tráfico.

Para la asignación de los diferentes tipos de tráfico a un valor PCP determinado se utiliza el mecanismo de prioridad estricta definido en IEEE 802.1Q. Este mecanismo permite realizar el mapeo entre las clases de tráfico (TC) y los PCP según el número de clases de tráfico que son admitidas por el puerto de salida del puente TSN. Es necesario ya que pueden darse situaciones en las que el número de clases de tráfico no coincidan con los 8 posibles PCP. De esta forma, si por ejemplo el puerto de salida del puente admite 8 clases de tráfico se asociará un valor PCP a cada TC, que poseerá su propia cola a la espera de ser transmitida por la interfaz. Pero si por el contrario el número de clases de tráfico es inferior a 8 entonces se deben fusionar varios PCP en una sola clase de tráfico y dicha cola poseerá tráfico de varios valores PCP. Este último caso es el proceso que se describe en la Figura 10.

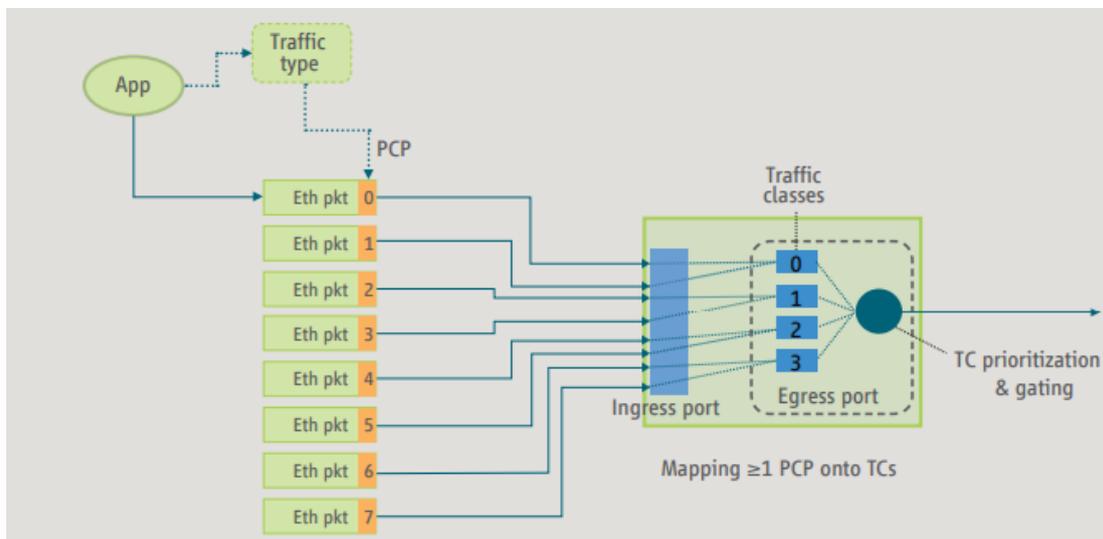


Figura 10: Mapeo entre los PCP y las clases de tráfico con IEEE 802.1Q [29]

Con el tráfico programado y la diferenciación en clases de tráfico se reduce considerablemente la interferencia que sufre los flujos TSN que circulan por la red. Sin embargo, existen ciertos tipos de tráfico que necesita una interferencia menor para cumplir con los requisitos de QoS. Para ello, se ha desarrollado el mecanismo de prevención de tramas (*Frame Preemption*) estandarizado en el IEEE 802.3br y en el IEEE 802.1Qbu. Este mecanismo permite que los paquetes de mayor prioridad se pueden adelantar al tráfico de menor prioridad para cumplir con sus requisitos de interferencia. Se puede utilizar junto con el tráfico programada para reducir aún más la interferencia.

Por otro lado, para conseguir una alta fiabilidad se utiliza la redundancia de los paquetes perteneciente a las comunicaciones más críticas. *Frame Replication and Elimination for Reliability (FRER)* definido en IEEE 802.1CB es el mecanismo utilizado por TSN para incrementar los paquetes que pertenecen a un determinado flujo. Con este mecanismo se garantiza que la red no va a sufrir una sobrecarga innecesaria ya que puede detectar los flujos duplicados y la combinación de ellos. De esta forma, se garantiza la transparencia de la redundancia para la aplicación y su realización dentro de la propia red. *FRER* implementa dos funciones principales cuyo funcionamiento se puede observar en la Figura 11:

- **Función de replicación:** se encarga de duplicar los paquetes y/o tramas y transmitirlos por dos o más rutas disjuntas. A cada una de las copias se les asigna el mismo número de secuencia para facilitar su posterior eliminación.
- **Función de eliminación:** se encarga de eliminar los paquetes y/o tramas adicionales que se reciben posteriormente a otra. Para saber si el paquete o trama recibido pertenece a un mensaje ya transmitido se debe examinar el número de secuencia asociado a cada paquete/trama.

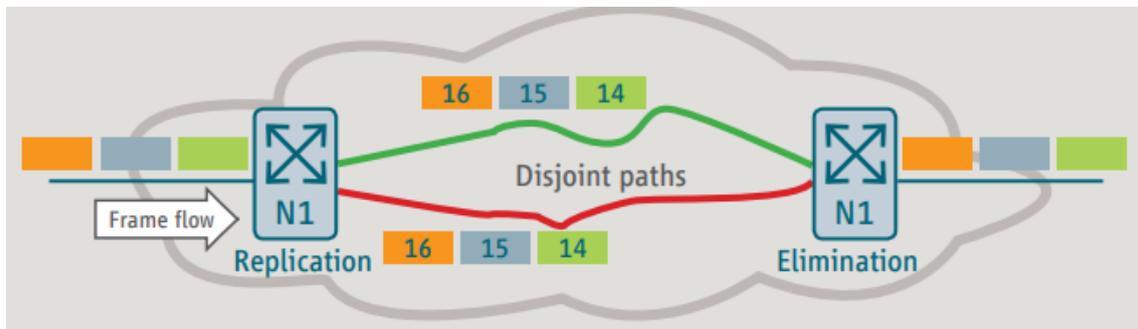


Figura 11: Funcionamiento del mecanismo FRER [29]

También se ha implementado el mecanismo de filtrado y vigilancia por flujo (*per-stream filtering and policing, PSFP*) estandarizado en IEEE 802.1Qci. PSFP permite identificar y gestionar el tráfico no conforme, así como el uso excesivo del ancho de banda intencionado o no, o la priorización incorrecta dentro de un intervalo de tiempo determinado. Este mecanismo implementa diferentes acciones de control como medidores de flujo que proporcionan vigilancia basada en los datos o puertas de flujos que proporciona vigilancia basada en el tiempo. La vigilancia basada en datos utiliza los medidores de flujo que se aplican a uno o más flujos TSN. Estos medidores proporcionan la tasa de información comprometida y la tasa de información excedente de los flujos para comprobar si excede la tasa de información de transmisión permitida. Por otro lado, la vigilancia basada en tiempo utiliza las puertas de flujo que requieren la sincronización de tiempo entre los puentes y las estaciones. Es necesario ya que las puertas son configuradas para que se abran sólo en instantes determinados de tiempo y aquellas tramas que llegan fuera de esa franja horaria son consideradas como interferencias o tramas no deseadas. Estas tramas son descartadas debido a que la puerta está cerrada y no permite el acceso de las tramas al puente.

## 2.2.1 TSN síncrono

Las redes TSN síncronas destacan porque todos los nodos que conforman la red presentan una misma referencia de tiempo, con el objetivo de garantizar su comportamiento determinista. Para ello, TSN emplea el protocolo de tiempo de precisión generalizado (*gPTP*, *generalized Precision Time Protocol*) especificado en el IEEE 802.1AS. Esta sincronización de reloj es necesaria para determinadas aplicaciones, así como para admitir tráfico programado TSN (*Scheduled Traffic*, IEEE 802.1Qbv) y para el uso de colas cíclicas (*Cyclic Queuing*, IEEE 802.1Qch). En esta sección se explica el procedimiento para la sincronización temporal de los nodos, los dominios de tiempo y el proceso de *scheduling*.

### 2.2.1.1 Proceso de sincronización temporal

Todos los nodos presentan una misma referencia de tiempo gracias a las compensaciones temporales que se realizan entre ellos para evitar que se produzcan retrasos y desvíos en la hora del reloj al asignarlo a un nodo. De esta forma, si los relojes presentan la misma referencia de tiempo es posible determinar el retardo que se produce extremo a extremo en toda la red para garantizar el requisito de latencia de una comunicación. Todos aquellos puentes TSN y estaciones finales TSN que comparten una misma referencia de tiempo conforman un dominio gPTP. El proceso de sincronización de tiempo consta de dos etapas:

- Establecimiento de la jerarquía Master-Slave:** Primeramente, se debe seleccionar la fuente de reloj que comparte su referencia de tiempo. El nodo seleccionado se denomina *Grand Master* (GM) y el reloj compartido puede ser universal o un reloj de trabajo basado en una escala de tiempo arbitraria. La información del reloj se distribuye desde el GM a todos los nodos del dominio con los mensajes del protocolo gPTP. La transmisión del reloj se realiza desde el nodo más cercano al nodo GM, denominado *Clock Master* (CM) hacia el siguiente nodo enlazado, denominado *Clock Slave* (CS). Todos los nodos que conforman el dominio gPTP actúan como CM y/o CS para la transmisión y la recepción de la referencia de tiempo, respectivamente. En la Figura 12 se representa un ejemplo del modo de actuación de los nodos de una red TSN en un dominio gPTP donde se comparte el reloj de uno de los puentes TSN o estación final.

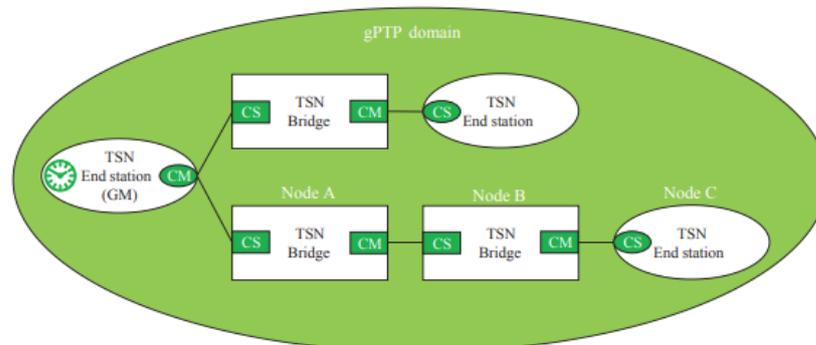


Figura 12: Ejemplo de un dominio gPTP donde se comparte un reloj según IEEE 802.1AS [16]

- Secuencia de mensajes gPTP de transmisión del reloj GM:** El protocolo gPTP calcula los retrasos que se produce en la transmisión del tiempo al tener en cuenta el tiempo de residencia de la trama dentro de cada puente, es decir, el tiempo consumido

para el procesamiento, la puesta en el buffer, y la transmisión desde los puertos de entrada a los de salida. Además, también considera la latencia de enlace de cada salto, es decir, el retardo de propagación que se produce entre dos puentes adyacentes. El cálculo de los retardos conlleva una secuencia de operaciones e intercambio de mensajes que se detallan a continuación para el Nodo B de la Figura 13 por ejemplo:

1. El *Clock Master* (CM), Nodo A, envía el *Sync Packet* (PDU gPTP) al *Clock Slave* (CS), Nodo B, que contiene una marca de tiempo de la hora de salida del mensaje del *Grand Master* (GM), según el reloj de referencia del GM, y el campo de corrección de retardo. El campo de corrección de retardo es el retardo acumulado de los anteriores puentes teniendo en cuenta el tiempo de residencia y propagación de cada puente. Este campo se emplea para calcular el reloj de referencia del CS que recibe el mensaje junto con la marca de tiempo de salida del mensaje del GM.
2. A continuación, se calcula el retardo de propagación del enlace que conecta directamente con el CM (Nodo A) y CS (Nodo B). Para ello, el CS envía el mensaje *Delay Request* al CM con el tiempo de salida de dicho paquete ( $t_1$ ) con el objetivo de calcular el tiempo de propagación. A continuación, el CM marca el tiempo de llegada del paquete anterior ( $t_2$ ) y lo envía junto con la marca de tiempo de salida ( $t_3$ ) del mensaje *Delay Response*. El CS recibe dicho mensaje y lo marca con su tiempo de llegada ( $t_4$ ). De esta forma, el CS puede determinar el tiempo de su reloj añadiendo el tiempo de propagación del enlace:

$$d = ((t_2 - t_1) + (t_3 - t_4)) / 2$$

En la Figura 13 se el intercambio de los mensajes anteriormente comentados entre el CM de un puente TSN y el CS de otro puente TSN, así como las marcas temporales.

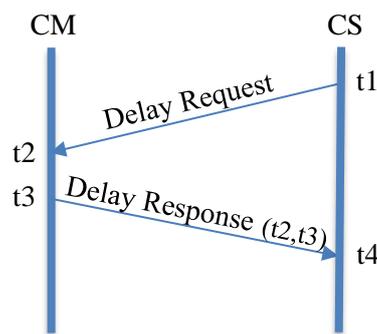


Figura 13: Intercambio de mensajes para calcular el retardo de propagación

3. Una vez obtenido el retardo de propagación, se actualiza la representación del reloj GM. Para ello, se incorpora el retardo de propagación calculado con anterioridad al retardo acumulado del campo de corrección y se suma este valor de retardo global a la marca del tiempo del GM.
4. Se calcula el tiempo de residencia del mensaje en dicho puente (Nodo B). El tiempo de residencia es el tiempo transcurrido desde que se recibe la PDU gPTP (puerto de entrada) en el puente hasta su transmisión (puerto de salida). Se

calcula de forma similar al proceso de obtención del retardo de propagación.

5. Tras la obtención del tiempo de residencia y del retardo de propagación se modifica el campo de corrección del mensaje *Sync Packet*. De modo que se suma ambos valores al retardo de tiempo acumulado incluido en el mensaje *Sync Packet* para la actualización de dicho campo.
6. Por último, se envía el mensaje *Sync Packet* con sus campos actualizados desde el CM del nodo B hasta los CSs que están directamente conectados (Nodo C).

Todo este proceso puede sufrir fluctuaciones debido a posibles retardos en los buffers y en el tiempo de propagación de los mensajes. Estos mensajes al ser periódicos se puede determinar su frecuencia de actualización para que el retardo que puede producir con el paso del tiempo no afecta a los relojes.

### **2.2.1.2 Dominios de tiempo**

La sincronización se puede realizar con un reloj de trabajo o universal. Para el reloj de trabajo, la desviación máxima entre el reloj de un nodo y el reloj de trabajo del GM debe estar entre 100 ns y 1  $\mu$ s. Un dominio de tiempo es el conjunto de nodos de la red que comparten el mismo reloj para el proceso de sincronización temporal, independientemente del tipo de reloj. Además, se pueden definir distintos dominios de trabajo redundantes para el mismo conjunto de dispositivos, logrando que si el GM falla el retardo experimentado sea mínimo al determinar un nuevo GM. Es decir, el tiempo de tolerancia a fallos sea mínimo. Los dominios de trabajo redundantes se pueden realizar a través de *cold standby* o *hot standby*. En ambos casos, un nodo en el dominio gPTP está preparado para asumir el papel del GM. En el caso del *cold standby*, el GM redundante espera hasta que el GM falle. Mientras que en *hot standby* el GM redundante comparte su reloj a través de las PDU gPTP incluso cuando el GM está activo, con el objetivo de realizar rápidamente el intercambio de roles cuando el GM falle, minimizando el tiempo de interrupción. Normalmente este reloj de trabajo se asocia a un conjunto de máquinas, celdas o líneas individuales.

Para la sincronización con el reloj universal, el reloj es proporcionado por una fuente externa o un propio nodo del sistema. Se permite una precisión más relajada entre el reloj del nodo y del GM (hasta 100 $\mu$ s), y normalmente se asocia a toda la planta o fábrica. Además, no son necesarios dominios redundantes por si falla el GM.

En la Figura 14 se representa una fábrica con diferentes dominios temporales, donde existe un dominio de tiempo global (se observa con el reloj y línea discontinua gris) para toda la fábrica y tres dominios de trabajo (el reloj verde, naranja y morado, y línea discontinua morada en la figura). También existe un dominio de tiempo TSN (el reloj y línea discontinua amarillas) que es otra variedad de dominio de tiempo. Este dominio se comparte entre aquellos nodos y puentes TSN de la red donde se transmite determinados tipos de tráfico que necesitan también la sincronización temporal de la aplicación, la red y la *scheduling*. Se emplea el mismo reloj para la sincronización de la aplicación como el acceso a la red con el objetivo de reducir el tiempo de sincronización.

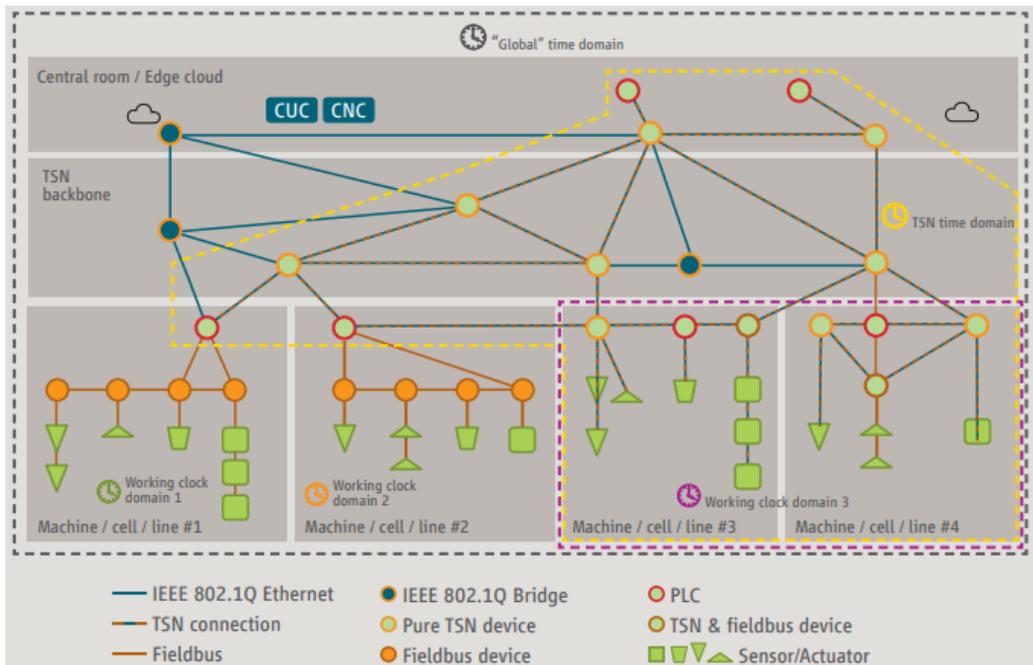


Figura 14: Varios dominios de sincronización temporal dentro de una fábrica [29]

### 2.2.1.3 Proceso de *scheduling*

TSN síncrono permite programar el tráfico (*Scheduled Traffic*, IEEE 802.1 Qbv) de las comunicaciones a través del mecanismo de colas cíclicas (*Cycling Queuing*, IEEE 802.1Qch). Esta técnica asignando instantes de tiempo para transmitir todas las tramas que pertenezcan a una determinada clase de tráfico. Es decir, TSN síncrono divide el tiempo en múltiples intervalos temporales que se repiten cíclicamente y dentro de cada uno de estos intervalos se transmiten los paquetes pertenecientes al conjunto de clases de tráfico establecido para el intervalo. Por lo que los paquetes que pertenecen a otra clase deben esperar en memorias intermedias hasta que se alcance su intervalo de tiempo (*slot*). En la Figura 15 se representa cómo el tiempo se divide en ciclos periódicos y estos a la vez en slots temporales con distintas duraciones, donde se transmite el tráfico asignado.

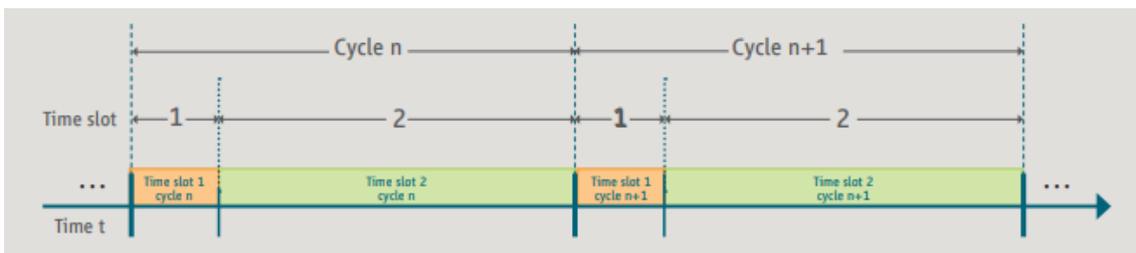


Figura 15: División del tiempo en intervalos temporales [29]

El tráfico que se transmite por las redes TSN síncrona presentan distintos requisitos de QoS, por tanto, es necesario agruparlo según sus prioridades para cumplir con las exigencias de retardo y *jitter*. Para ello, dependiendo de los requerimientos de dicha trama se le asigna un valor PCP concreto. Una vez asignado el valor PCP a la trama, esta se agrupa en colas intermedias en los puertos de salida del puente dependiendo de la clase de tráfico asociado a dicho PCP. Es decir, se utiliza el mecanismo de prioridad estricta descrito en la Sección 2.2 TSN. La trama tras encolarse en una cola de tráfico se transmite en los slots temporales según el TC asociado a dicha trama. Para llevar a cabo este proceso TSN síncrono define el *scheduler Time Aware Shaper (TAS)*

especificado en IEEE 802.1Qbv. Este *scheduler* implementa un mecanismo de control de puerta que habilitar o deshabilitar la puerta de salida a la interfaz de red de cada una de las colas de tráfico. Este proceso está determinado por una lista que contienen las puertas que deben abrirse y cerrarse en cada instante temporal, como se observa en la Figura 16. De esta forma, es posible que en un intervalo de tiempo se puedan transmitir tráfico de varias clases. Para su correcto funcionamiento y que no se produzcan fallos en la transmisión es necesario la correcta sincronización temporal entre los nodos de la red.

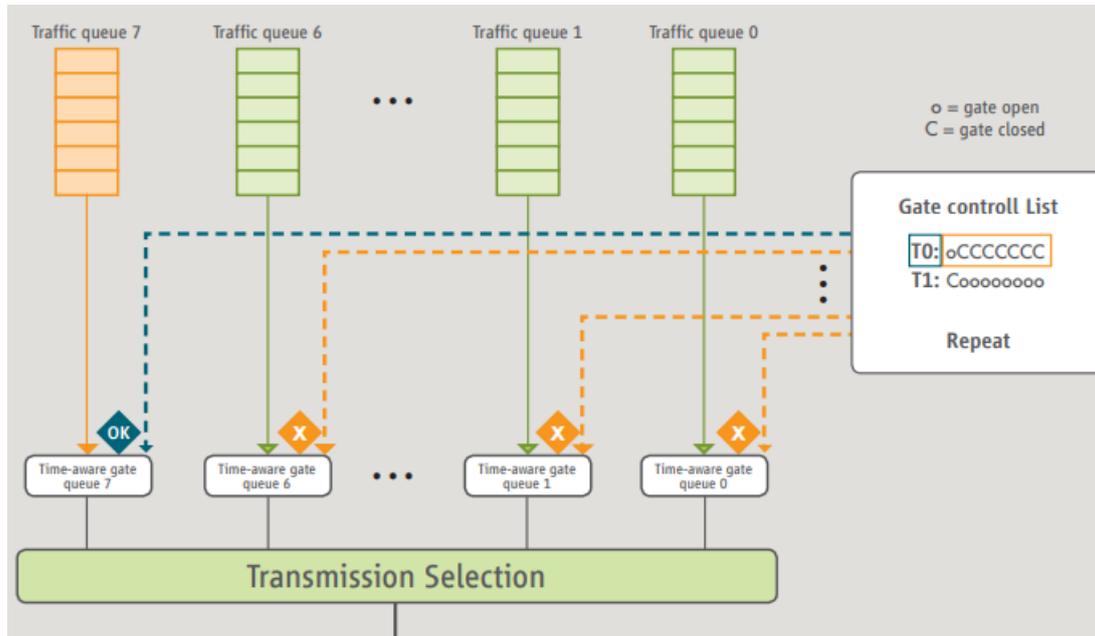


Figura 16: Ejemplo del funcionamiento del mecanismo de control de puertas [29]

En la Figura 17 se representa una descripción a alto nivel de la secuencia de operaciones que sufre una trama TSN antes de ser transmitida por el puente TSN. Los procesos que se muestran son la clasificación del tráfico, el *scheduler TAS* y el *preemption*.

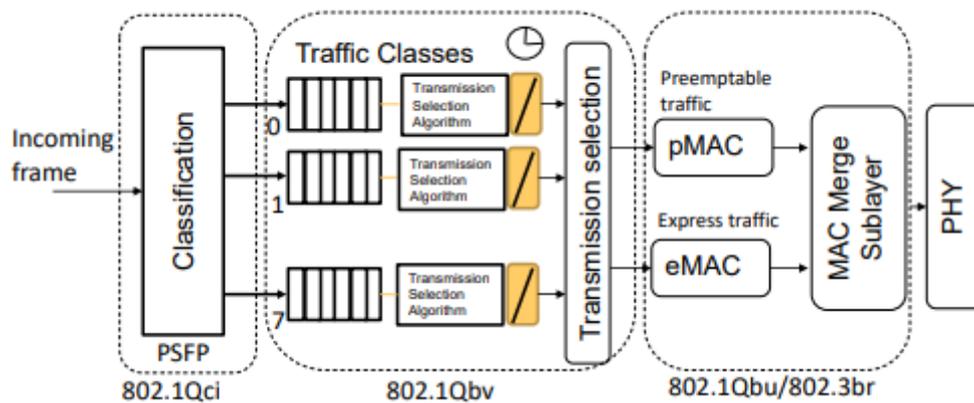


Figura 17: Procesamiento de tramas considerando IEEE 802.1Qci, IEEE 802.1Qbv, IEEE 802.1Qbu y IEEE 802.3br [4]

### **2.2.2 TSN asíncrono**

En TSN asíncrono no son necesarios los mecanismos de sincronización temporal (IEEE 802.1AS) ni los algoritmos de *scheduler* de tráfico TSN síncrono (IEEE 802.1Qch y IEEE 802.1Qbv). Porque los nodos de las redes TSN asíncronas no necesitan una misma referencia temporal en sus relojes para cumplir con los requisitos de QoS demandados por los flujos. Además, las redes TSN síncronas presentan grandes debilidades debido a la necesidad de coordinación temporal entre sus nodos, que dificulta la escalabilidad de la red, y al uso de slots de tiempo reservado para los flujos, que provoca una utilización pobre de la capacidad de los enlaces. Por tanto, las redes TSN asíncronas implementan otro *scheduler* que no necesita una referencia de tiempo común entre todos los puentes TSN y mejora la escalabilidad y utilización de los enlaces.

El *scheduler* que utiliza las redes TSN asíncronas es *Asynchronous Traffic Shaper (ATS)* estandarizado en el IEEE 802.1Qcr. Este mecanismo está basado en *Urgency-Based Shaper (UBS)* propuesto por los investigadores Specht y Samii [5]. El ATS desarrolla un método asíncrono para el procesamiento de las tramas TSN en los puertos de salida del puente TSN. Primeramente, se explica el funcionamiento de UBS para comprender posteriormente las diferentes etapas de la que se compone el *scheduler* ATS, ya que algunas de estas etapas desarrollan procesos similares a los de UBS.

En la Figura 18 se muestra la arquitectura y el proceso de UBS. En este caso, para simplificar la representación solo se muestra un puerto de salida del puente TSN pero esta misma arquitectura se repite en todos los puertos de salida del puente TSN para UBS. El conformador UBS consta de dos etapas con una serie de colas que siguen una disciplina FCFS (*First Come, First Served*): una con colas conformadas para las configuraciones intercaladas y otra con colas dependiendo de la prioridad.

- **Conformación intercalada (*Interleaved shaping*):** En esta etapa se puede utilizar una única cola (cola conformada) para la regulación del tráfico de un conjunto de flujos con requisitos de QoS propios. Por tanto, para determinar si la trama se puede transmitir o no solamente se comprueba la admisión de la trama de cabecera de línea (*head-of-line*, HOL). Es decir, se comprueba si la trama HOL es elegible para la transmisión según las restricciones de regulación de dicho flujo. Si es así, se produce la liberación de la trama para su transmisión a la siguiente etapa. En determinadas investigaciones se ha determinado que la utilización del conformado intercalado no incrementa la latencia en el peor de los casos para UBS [5][6]. Además, UBS soporta restricciones del conformador de cubo con escape con el objetivo de imponer una tasa de datos comprometida y un tamaño de ráfaga para cada flujo. Es decir, los flujos que se transmiten están limitados a un tasa superior de la forma:

$$A_f(t) \leq r_f * t + b_f$$

Donde  $A_f$  corresponde a la cantidad de datos que se han transmitido hasta un instante de tiempo  $t$  para la clase de flujo  $f$ , mientras que  $r_f$  es la tasa de datos comprometida y  $b_f$  el tamaño de ráfaga comprometida [8].

- **Colas de prioridad (*Priority Queues*):** Esta etapa está formada por una cola con

disciplina FCFS para cada uno de los niveles de prioridad considerados en el *scheduler*. Cada cola de una determinada prioridad agrupa las salidas de las colas conformadas asociada a dicho nivel de prioridad según indica su TC. Por tanto, una cola de prioridad P puede presentar tráfico procedente de diferentes colas conformadas. Para la transmisión de los flujos de cada una de las colas se sigue un mecanismo de selección de prioridad estricta. Este algoritmo da mayor preferencia a las colas de mayor prioridad para el acceso al medio compartido que a las colas con menor prioridad. Es decir, a pesar de que una trama de prioridad P esté lista para ser transmitida debe esperar a la transmisión de las tramas que se encuentren en las colas con prioridad superior a P.

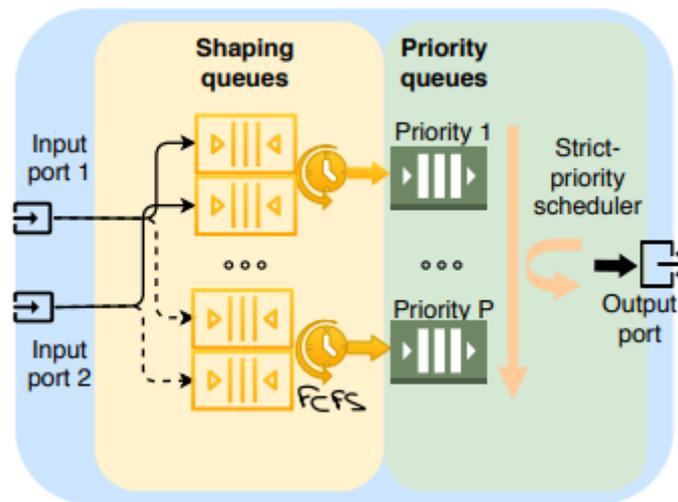


Figura 18: Arquitectura y funcionamiento de UBS [17]

La asignación de un flujo a un *scheduler* ATS determinado depende de dos factores importantes: los flujos que procesa la cola conformada y los niveles de prioridad asociados a las colas de prioridad. Sin embargo, estas decisiones se determinan a través de las siguientes tres reglas [17]:

- Cada una de las colas conformadas está asociada solamente a un puerto de entrada (regla QAR1).
- Cada cola conformada está asociada con un nivel de prioridad en el salto/puente anterior (regla QAR2).
- Cada cola conformada está asociada con un nivel de prioridad interno (regla QAR3).

Al emplear estas reglas se consiguen beneficios para la red TSN asíncrono. Es decir, QoS determinista se proporciona con las reglas QAR2 y QAR3 mientras que la regla QAR1 permite aislar los nodos de la red para evitar que tráfico no conformado sobrecargue la red. Además, el uso de estas reglas permite calcular el número de colas conformadas necesarias para implementar un nivel de prioridad. Por ejemplo, dado un ATS con un número  $U$  de puertos de entrada y donde cada puerto de entrada  $u \in [1, U]$  recibe tráfico asociado con un nivel de prioridad  $P_u$  del nodo adyacente. Entonces las colas conformadas necesarias para el correcto funcionamiento con  $P$  niveles de prioridad internos y sin restricción de asignación de prioridad como mínimo son  $Q = P * \sum_{u=1}^U P_u$ .

Por tanto, la estructura global del ATS se muestra en la Figura 19, donde se produce la asignación de los flujos al *scheduler* ATS adecuado y se implementa el mecanismo UBS. Este procedimiento consta de cuatro fases, donde las tres primeras se corresponden con la primera etapa de UBS y la última fase del ATS con la segunda etapa de UBS. Las fases son:

- **Filtrado de flujos (*Stream filtering*):** En esta etapa se produce la clasificación y el filtrado del ATS según el flujo de entrada. Cada flujo es asignado a un filtro según su prioridad y su identificador de conexión. Puede producirse que alguna de las tramas exceda la unidad máxima de datos de servicio (*Service Data Unit, SDU*) por lo que los filtros tienen la capacidad de bloquear dicha trama. Además, asignan a cada flujo la compuerta de la siguiente etapa y el *scheduler* ATS asociados a ese filtro.
- **Compuerta de flujo (*Stream Gating*):** En esta etapa se asigna el valor de prioridad interno (*Internal Priority Value, IPV*) a un flujo anulando su prioridad real, indicado por su PCP, para el proceso de reenvío en la etapa de encolado y transmisión. Esto se produce para facilitar el cumplimiento de los requisitos E2E de los flujos, ya que permite dar mayor prioridad aquellos flujos con mayores restricciones de retardo.
- **Planificación ATS (*ATS scheduling*):** Esta fase implementa el proceso descrito en la primera etapa de UBS. Es decir, los flujos se acumulan en colas conformadas hasta que la trama HOL pueda ser transmitida a la siguiente etapa según las restricciones de regulación del flujo.
- **Encolado y transmisión (*Queuing and Transmission*):** En esta fase los flujos salientes de las colas conformadas se colocan en los *buffers* dependiendo de la clase de tráfico (TC) asociada, de manera similar a la segunda etapa de UBS. Por lo que es necesario una tabla en los puertos de salida que mapee los niveles de prioridad de los flujos a las clases de tráfico para asignar los flujos a las colas TC correcta. Si la asignación de IPV está habilitada entonces el mapeo de los flujos se realizan con este valor en lugar del nivel de prioridad real de la trama. Una vez que están en las colas se transmiten al medio compartido a través de un *scheduler* de prioridad estricta.

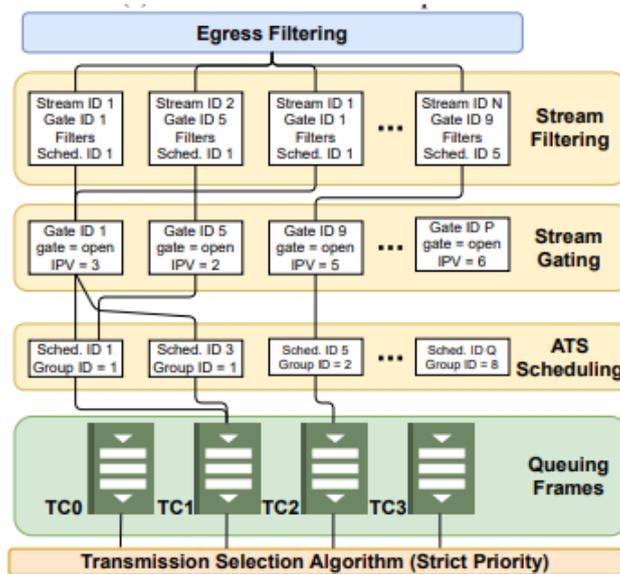


Figura 19: Arquitectura y funcionamiento del scheduler ATS [17]

### **2.2.3 Ventajas e inconvenientes de TSN síncrono y asíncrono**

La mayoría de los puentes TSN implementados en el estándar TSN son síncronos ya que necesitan de una referencia de tiempo común entre los nodos de la red para cumplir con los requisitos de latencia y *jitter* exigidos. Sin embargo, como se ha comentado anteriormente, esto provoca que la escalabilidad sea más costosa y la utilización de la capacidad de los enlaces mínima debido a la reserva de los slots temporales. Para solucionar estos problemas de escalabilidad y utilización se implementa las redes TSN asíncronas, donde no es necesario la sincronización temporal. Pero provoca un incremento del retardo que sufren los flujos TSN, aunque siguen cumpliendo con las restricciones exigidas. En la Tabla 3 se compara más en profundidad las ventajas e inconvenientes que presentan las redes TSN síncronas y asíncronas.

<b>Tipo de red</b>	<b>Redes TSN síncronas</b>	<b>Redes TSN asíncronas</b>
<b>Ventajas</b>	<ul style="list-style-type: none"> <li>- Latencia inferior.</li> <li>- Adaptable a tráfico determinístico periódico.</li> </ul>	<ul style="list-style-type: none"> <li>- Mayor escalabilidad que las redes síncronas.</li> <li>- Mayor utilización de los enlaces que las redes síncronas.</li> <li>- Menor complejidad de configuración que las redes síncronas.</li> <li>- Adaptable al tráfico determinístico aperiódico/espórádico.</li> </ul>
<b>Desventajas</b>	<ul style="list-style-type: none"> <li>- No adaptable a tráfico determinístico aperiódico.</li> <li>- Emplea tecnología más cara.</li> <li>- Mayor dificultad de configuración</li> <li>- Menor utilización de enlace</li> <li>- Problemas escalabilidad debido a la sincronización temporal</li> </ul>	<ul style="list-style-type: none"> <li>- Mayor latencia que las redes síncronas.</li> <li>- Más caro que las redes asíncronas tradicionales.</li> <li>- No adaptable a tráfico determinista periódico.</li> <li>- Menor utilización de enlace que las redes asíncronas tradicionales.</li> <li>- Mayor complejidad de configuración que las redes asíncronas tradicionales.</li> </ul>

*Tabla 3: Comparación entre redes TSN síncronas y asíncronas [7]*

## **2.3 Integración de TSN y 5G**

Las tecnologías que actualmente se utilizan en 5G no presentan el retardo suficiente para ser usadas en las Industrias 4.0, y tampoco permiten controlar el retardo de igual manera que se puede realizar con la tecnología TSN. Para ello, en el 3GPP *Release 16*, se incluye una propia integración entre la red de acceso radio 5G y TSN síncrono para dar soporte a los procesos industriales y a la automatización, reduciéndose el gasto de cables y permitiendo una mayor flexibilidad. Por tanto, en este apartado se detalla la arquitectura desarrollada por 5GACIA [29] basándose en el 3GPP *Release 16*.

### **2.3.1 Arquitectura de integración de TSN y 5G según el 3GPP**

Los casos de usos que se desea implementan presentan características de latencia que la arquitectura 5G no puede proporcionar, por lo que se crea un modelo de forma que el sistema 5G actúe como un puente hacia la tecnología TSN síncrona. Es decir, el sistema 5G actúa como un

puente virtual de la red TSN síncrona, proporcionando conectividad en el plano de control (*Control Plane*, CP) y puertos TSN en el plano de usuario (*User Plane*, UP). De esta forma, el sistema 5G puede conectar a uno o dos puentes TSN síncronos con los dispositivos finales de manera transparente para la red TSN, es decir, aparentando ser otro puente TSN. Para conseguirlo, se introduce unos traductores TSN en los bordes del sistema 5G que se encargan de permitir su conexión con la red TSN síncrona, el reenvío del tráfico de datos en el UP y la configuración de los puentes TSN en el CP. Existen 3 traductores TSN (*TSN Translator*, TT):

- Uno en el plano de control a través de una función de aplicación TSN (*Application Function*, AF).
- Otro en el lado UE (*User Equipemnt*) mediante un TT del lado del dispositivo (*Device-side TT*, DS-TT).
- El último en el lado de la función del plano de usuario (*User Plane Function*, UPF) mediante un TT del lado de la red (*Network-side TT*, NW-TT).

Aclarar que el plano de usuario, donde se produce el reenvío del tráfico, está constituido principalmente por el equipo de usuario (UE), el punto de acceso radio (RAN) y la función de plano de usuario (UPF). El UE es el componente que permite la comunicación con el RAN. El RAN es el dispositivo que actúa como punto de acceso inalámbrico al puente TSN, denominada en las redes 5G como *Next Generation Node B* (gNB). El gNB puede comunicarse con uno o varios UE. Por último, el UPF permite el establecimiento de la comunicación entre el gNB y el traductor del lado de la red (NW-TT).

En el UP, DS-TT y NW-TT actúan como puertos Ethernet de entrada y salida en el puente lógico 5GS TSN. Estos traductores están conectados al UE y UPF, respectivamente y su funcionalidad es traducir todos los parámetros necesarios para coordinar la comunicación entre el sistema 5G y la red TSN. Es decir, se encargan del mapeo de los requisitos de los flujos TSN a los requisitos del sistema 5G según se han establecidos por el TSN AF. En la Figura 20 se representa la arquitectura detallada en el plano de usuario. También se muestra como los puentes TSN comunican sus propiedades y los requisitos de los flujos al CNC para que este la transmita al traductor TSN AF. Con esta información el TSN AF puede establecer una correcta configuración del puente lógico 5GS TSN según las restricciones de QoS de los flujos.

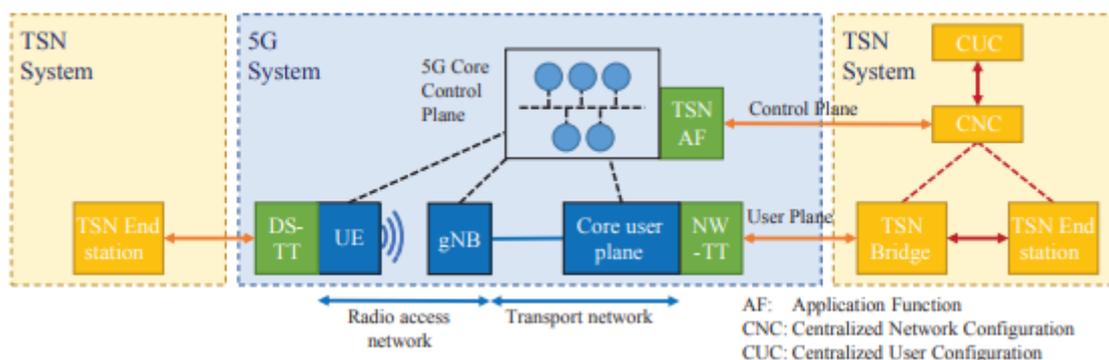


Figura 20: Arquitectura de integración TSN-5G del plano de usuario [15]

Cada uno de los puertos DS-TT tiene asociado una sesión de unidad de datos de protocolo (*Protocol Data Unit*, PDU) específica de los sistemas 5G, mientras que cada uno de los puertos NW-TT está asociado a un puerto físico en la UPF. Además, puede producirse que dentro de un mismo sistema 5G se creen distintos puentes virtuales TSN como se muestra en la Figura 21. Es

posible ya que un puente virtual está formado principalmente por una UPF y por todos los UE que se conecten a ella, y como en un sistema 5G puede existir varias UPF se puede crear varios puentes virtuales. Todas las sesiones PDU que se han formado en los puertos DS-TT y están conectados a una UPF determinada constituyen un grupo y pertenecen a un mismo puente virtual. Por tanto, a partir de un UE se pueden crear múltiples sesiones PDU a diferentes UPF con el objetivo de crear transmisiones redundantes o aislar el tráfico. De esta forma, el UE que presenta múltiples sesiones PDU a varios UPF está compartido entre varios puentes virtual, es decir, estaría compartido con el mismo número de puentes virtuales que UPF diferentes a las que está conectado cada sesión. Cada uno de los puertos DS-TT asignado a una sesión PDU pertenece a un puente virtual.

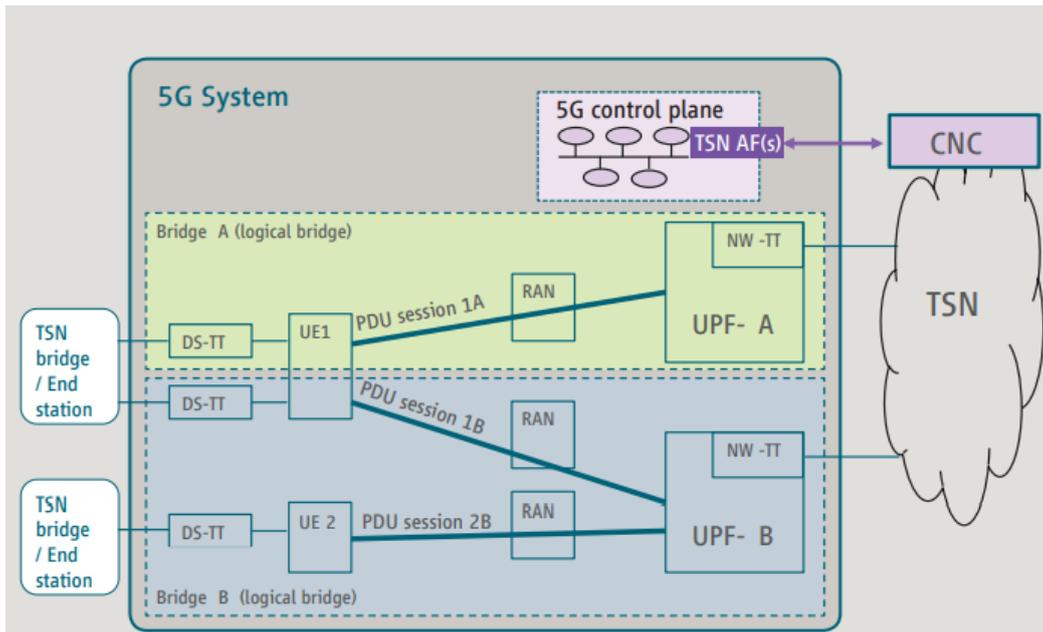


Figura 21: Arquitectura de integración TSN-5G con más de un puente virtual TSN [29]

En el plano de control, se establece el intercambio de los datos de control entre el puente lógico 5GS TSN y los puentes TSN, ya que el CP se encarga de la gestión de las conexiones, las políticas QoS, la autenticación y el resto de las funciones relacionadas con la gestión. Para ello, además de introducir el traductor TSN AF el puente 5G admite todas las funciones de red que conforman el 5G Core (5GC):

- **Función de gestión de acceso y movilidad (*Access and Mobility Management Function, AMF*):** se encarga de manejar la movilidad de las comunicaciones (*handover*) que se producen entre los gNB y de recibir la información relacionada con las sesiones 5G, como la autenticación y autorización de los accesos.
- **Sistema de gestión de red (*Network Management System, NMS*):** es la base de datos empleada en la gestión de red. Contiene toda la información de gestión incluyendo los parámetros de configuración de la red 5G, información sobre los flujos o almacenamiento de los KPIs (*Key Performance Indicators*), etc. Por consiguiente, se encarga de controlar toda la información relevante para el establecimiento de las conexiones extremo a extremo mediante el uso de las interfaces del sistema 5G, que previamente han sido comunicadas al CNC mediante el TSN AF. Es decir, configurar los nodos del sistema 5G para la comunicación extremo a extremo y su gestión.

Además, permite implementar la funcionalidad de *network slicing* en el sistema 5G.

- **Función de gestión de sesión (*Session Management Function, SMF*):** crea los canales de comunicación entre el CP y UP. Además, gestiona las sesiones de PDU, es decir, crea, actualiza y borra las PDUs y proporciona la dirección MAC de los traductores para cada sesión PDU al TSN AF a través del PCF.
- **Función de control de políticas (*Policy Control Function, PCF*):** se comunica con el TSN AF para recibir información sobre la QoS exigidas. Con esta información mapea los parámetros de QoS TSN a los valores de 5QI. Al igual se ha explicado en el anteriormente en la Sección 2.1 5G, el 5QI contiene ciertas características de QoS de 5G como la prioridad, el PDB, la PER y el MDBV.
- **Gestión de datos unificada (*Unified Data Management, UDM*):** presenta una funcionalidad similar al HSS (*Home Subscriber Service*) de las redes 4G. Almacena las credenciales de autenticación e información de los usuarios, como los componentes que están conectados de la red, los datos e información personalizada.
- **Función de exposición de red (*Network Exposure Function, NEF*):** proporciona una conexión segura entre los servicios 5G.
- En la Figura 22 se muestra la arquitectura detallada de integración del puente lógico 5GS TSN con puentes TSN tanto del plano de control como del plano de usuario.

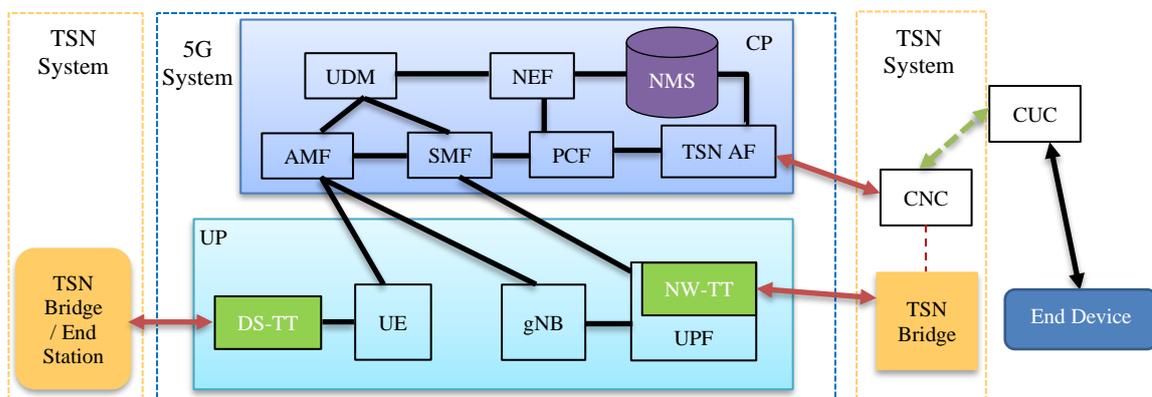


Figura 22: Arquitectura de integración TSN y 5G

### 2.3.2 Proceso de interacción entre el plano de control de 5GS y el CNC de los puentes TSN

La función de aplicación (AF) TSN ejerce un papel relevante en la interacción entre el plano de control del sistema 5G y TSN. Este traductor funciona como un “proxy” entre el CNC de una red TSN que utiliza las interfaces de gestión de red estandarizadas y el sistema 5G. De esta forma, no se produce la exposición a toda la red de detalles de configuración específicos para 5G. Las interacciones entre el CNC y el TSN AF se pueden resumir en tres fases, como se observa en la Figura 23:

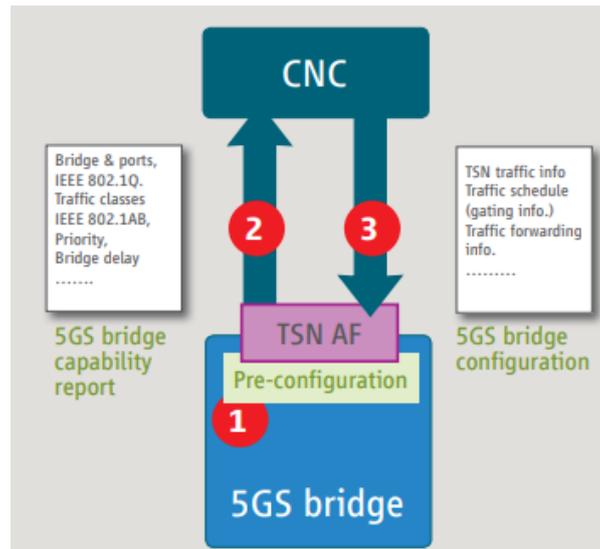


Figura 23: Interacción en el plano de control entre TSN y 5G [29]

- **Fase 1. Preconfiguración:** Se configura la información del puente 5G, es decir, se determina el ID del puente y los números de puertos que posee los puertos NW-TT, que pueden ser configurados con el UPF. Por otro lado, el TSN AF debe poseer una tabla de mapeo de QoS, donde se enumere las clases de tráfico posible y su relación con los perfiles de QoS 5G que están preconfigurados. Esta tabla también debe tener el retardo que se produce entre el UE/DS-TT y el UPF/NW-TT del puente 5G y los niveles de prioridad.
- **Fase 2. Informe sobre las capacidades del puente del sistema 5G.** El sistema 5G envía al TSN AF tanto la información que se ha preconfigurado como información de gestión de los puertos y los puentes durante la formación de las sesiones de PDUs. Asimismo, la SMF proporciona la información relacionada con la sesión de la PDU al TSN AF mediante el PCF. Con toda la información que obtenida el TSN AF genera un informe sobre el puente 5G para el CNC.

El CNC también recibe información sobre los múltiples puentes virtuales que se han formado en el sistema 5G. Para ello, el sistema 5G notifica al TSN AF sobre información específica del puente como su nombre, dirección, la cantidad de puertos e información del propio puerto, como el número y la dirección. Así como si los puertos están ubicados en el DS-TT o en el NW-TT.

Además, el TSN AF informa sobre la topología obtenida con el protocolo de descubrimiento de capa de enlace (LLDP). Así como de las determinadas capacidades del sistema 5G, como las demoras máximas garantizadas entre los múltiples puertos, que permiten al CNC programar adecuadamente el tráfico extremo a extremo. A igual que comunica las clases de tráfico que son admitidas por el sistema 5G para llevar a cabo la priorización del tráfico.

- **Fase 3. Configuración del puente.** En este caso es el sistema 5G el que recibe la información de configuración de TSN a través de TSN AF para configurar el puente 5G. Los datos transmitidos incluyen información de programación de los tráficos como se especifica en el IEEE 802.1Qbv, información de PSFP como se indica en el

IEEE 802.1Qci e información de reenvío de tráfico. Con los últimos datos el TSN AF puede determinar los puertos de entrada y salida para un determinado flujo.

Desde el punto de vista del TSN AF, un puente 5G TSN presenta un solo NW-TT dentro de una UPF, pero puede tener múltiples puertos de reenvío de tráfico. Por tanto, para transmitir correctamente los flujos de tráfico el NW-TT debe reenviarlo por el puerto de salida apropiado según se indica en la información de reenvío recibida en el TSN AF. Mientras que, para el reenvío del flujo en el lado del UE, el TSN AF selecciona el UE cuyo tráfico se va a encaminar a través de la dirección MAC del DS-TT utilizada en la sesión PDU.

Una vez que el tráfico TSN alcanza el puente 5G, este tráfico debe ser asignado a los flujos de QoS 5G adecuados en sus correspondientes PDU con la adecuada configuración de QoS. Para ello, el sistema 5G recibe la información de QoS del tráfico TSN a través del TSN AF desde el CNC. La identificación del perfil QoS 5GS adecuado para dicho tráfico se consigue usando la tabla de mapeo determinada en la fase 1 de preconfiguración. De esta forma, el sistema 5G utiliza este perfil para determina el flujo QoS 5G que entregue el tráfico TSN desde el puerto de entrada al puerto de salida del puente 5G. Como se puede observar en la Figura 24.

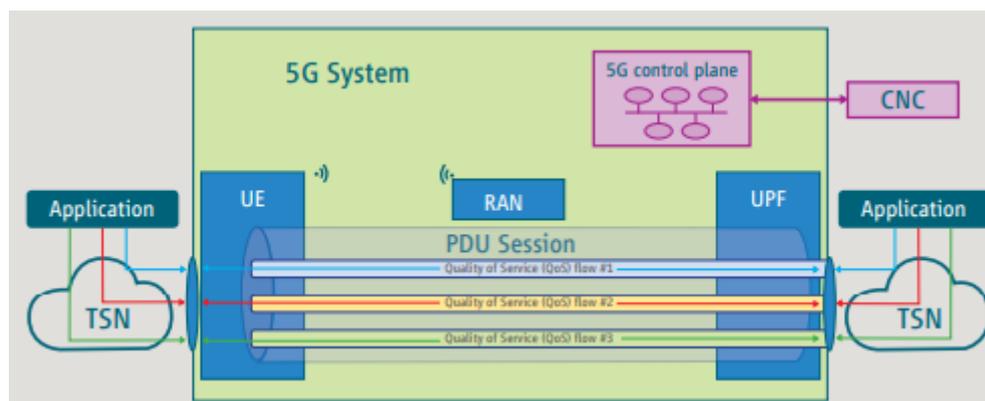


Figura 24: Mapeo de QoS entre tráfico TSN y flujos QoS 5G [29]

Los diferentes parámetros que caracterizan a un flujo QoS 5G, que se encuentran detallados en la Sección 2.1 5G, se determinan a partir de la información de configuración recibida por TSN AF. Los valores MFBR y GFBR se derivan de la información de PSFP recibida y para el ARP se establece un valor preconfigurado. Mientras que para las características que contiene el 5QI el 3GPP ha definido una serie de valores estandarizados. Pero puede conseguirse una mejor adaptación al tipo de tráfico TSN si, por ejemplo, se asigna un determinado nivel de prioridad 5QI según el valor del PCP especificado en el tráfico TSN.

Finalmente, con la información de PSFP recibida del CNC el TSN AF calcula la información de asistencia de comunicación sensible al tiempo (*Time-Sensitive Communication Assistance Information*, TSCAI). El TSCAI es proporcionado al RAN 5G para configurar la conectividad RAN y dar un soporte eficiente para los flujos TSN.

El puente 5GS debe implementar todas las funciones que caracterizan a la tecnología TSN síncrona, sin embargo, no es necesario el uso de todas ellas para transmitir las comunicaciones

TSN a menos que sea necesario por los requisitos requeridos. Asimismo, el puente 5G puede funcionar como un puente TSN sin capacidades específicas.

### **2.3.3 Proceso de sincronización temporal**

En la integración de TSN síncrono con 5G también es necesario la sincronización de una referencia de tiempo común. Entonces, el sistema TSN síncrono presenta su propio reloj de referencia mientras que el sistema 5G el suyo, aunque puede suceder que el reloj del puente 5G sea también el utilizado por el puente TSN para evitar tener que trabajar con dos relojes distintos que complican el proceso de sincronización. Según se ha calculado en [27] el *budget* de sincronización que se puede producir en el sistema 5G para un dominio de tiempo global no debe exceder los 900 ns. Existen dos procesos de sincronización convergentes:

- **Sincronización 5GS:** el nodo GM (*Grand Master*) proporciona el reloj de referencia para la sincronización interna del sistema 5G. Por tanto, todos los nodos del sistema presentan el mismo reloj incluso los traductores. Este reloj de referencia se transmite a través de la red de transporte hacia el UPF y NW-TT con el protocolo PTP y a través del enlace inalámbrico al UE y DS-TT. El gNB al adquirir el tiempo de referencia puede modificar su valor gracias al tiempo de propagación de las tramas que utiliza 5G *New Radio* (NR), cuya estructura permite calcular una referencia de tiempo precisa. Tras la modificación del tiempo de referencia el gNB lo transmite a los UE en un bloque de información del sistema (*System Information Block*, SIB) o en un mensaje de control de recursos de radio (*Radio Resource Control*, RRC). Al usar el enlace inalámbrico se producen fluctuaciones en el tiempo de transmisión que pueden variar en los mensajes transmitidos continuos. Este *jitter* se evita al realizarse la reserva de recursos en la RAN 5G, ya que los mensajes gPTP son periódicos y se consideran como mensajes de máxima prioridad, por lo que se reservan recursos para su transmisión.
- **Sincronización dominio TSN:** proporciona el reloj de referencia para los dispositivos TSN. El proceso sería similar al descrito anteriormente. Sin embargo, ahora se debe tener en cuenta el retardo introducido por el sistema 5G. Para ello, cuando el mensaje *PDU gPTP* procedente de la red TSN llega al NW-TT este genera una marca de tiempo de entrada (Tsi) basado en el reloj interno 5GS e incrusta la marca de tiempo en el mensaje gPTP como sufijo. La UPF reenvía el mensaje gPTP al UE a través del plano de usuario encapsulado en una sesión PDU. El UE recibe el mensaje gPTP y lo reenvía al DS-TT. El DS-TT crea una marca de tiempo de salida del mensaje (Tse) basado en el reloj interno del 5GS. Para calcular el retardo que se ha producido en la transmisión del mensaje dentro del 5GS (tiempo de residencia) se realiza la diferencia entre Tsi y Tse. Por último, el DS-TT modifica la información de temporización de TSN recibida en los mensajes de gPTP para añadir el tiempo de residencia y envía la PDU gPTP al siguiente puente TSN o estación final.

El proceso descrito se representa en la Figura 25. Ambos procesos de sincronización pueden considerarse como independiente entre ellos, lo que aporta flexibilidad en la implementación de la sincronización del tiempo. Así como la posibilidad de introducir nuevas mejoras en alguno de los sistemas sin tener que modificar por completo toda la sincronización de la red. De esta forma, por ejemplo, si se introduce un sistema 5G a una red TSN fija la sincronización del tiempo TSN

no es afectada por la introducción de 5G.

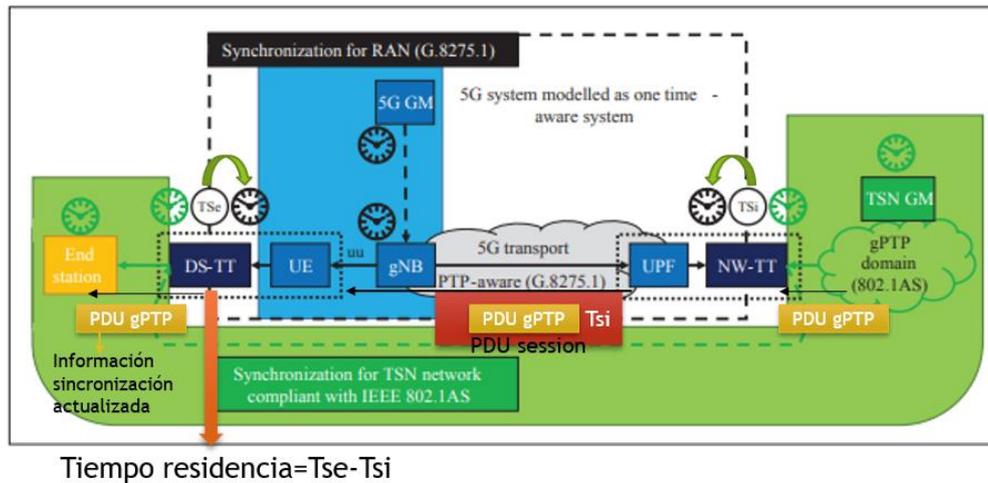


Figura 25: Sincronización temporal de la integración entre TSN y 5G

### 2.3.4 Dominios de tiempo

Un sistema 5G también admite múltiples dominios de tiempo, tanto dominios con relojes de trabajo como de tiempo global. Mientras que el número de dominios que una industrial TSN puede soportar depende del tipo de dispositivo:

- **Dispositivos restringidos (clase B):** tienen al menos dos dominios de gPTP, un dominio de reloj de trabajo y un dominio de reloj universal.
- **Dispositivos de altas características (clase A):** tienen al menos dos de cada dominio de tiempo (es decir, 4 en total).

En el Rel-16, se admiten hasta 32 dominios de reloj de trabajo diferentes en los sistemas 5G. Cada dominio de tiempo se identifica por un número específico de dominio en las PDU gPTP. De esta forma, se permite calcular el tiempo de residencia de los diferentes dominios en el sistema 5G. El procedimiento es igual al proceso de sincronización de dominio TSN descrito anteriormente. El NW-TT genera un Tsi para cada mensaje gPTP indicando el número de dominio específico al que pertenece dicho mensaje. El UE recibe los mensajes gPTP y los reenvía al DS-TT. El DS-TT recibe en el mensaje gPTP tanto la información de temporización del reloj gPTP original como el Tsi correspondiente según el dominio de trabajo. Cuando recibe el mensaje gPTP, el DS-TT marca el tiempo de llegada (Tse) para cada uno de los dominios de trabajo y calcula el tiempo de residencia de dicho mensaje gPTP para ese dominio de trabajo. De nuevo, el mensaje gPTP se regenera en el DS-TT con la modificación de la información de temporización de TSN incluyendo el tiempo de residencia del sistema 5G y se envía la PDU gPTP al siguiente puente TSN o a la estación final. Este proceso se repite para el resto de los mensajes gPTP con dominio de trabajo distinto. Por tanto, una estación final puede acceder a la información temporal según el número de dominio del mensaje gPTP. Ambas marcas temporales (Tsi y Tse) se han referenciado según el reloj del sistema 5G con el que se han sincronizado las entidades NW-TT y DS-TT. En la Figura 26 se muestra la interacción entre del sistema 5G y los puentes TSN cuando se definen varios dominios de trabajo.

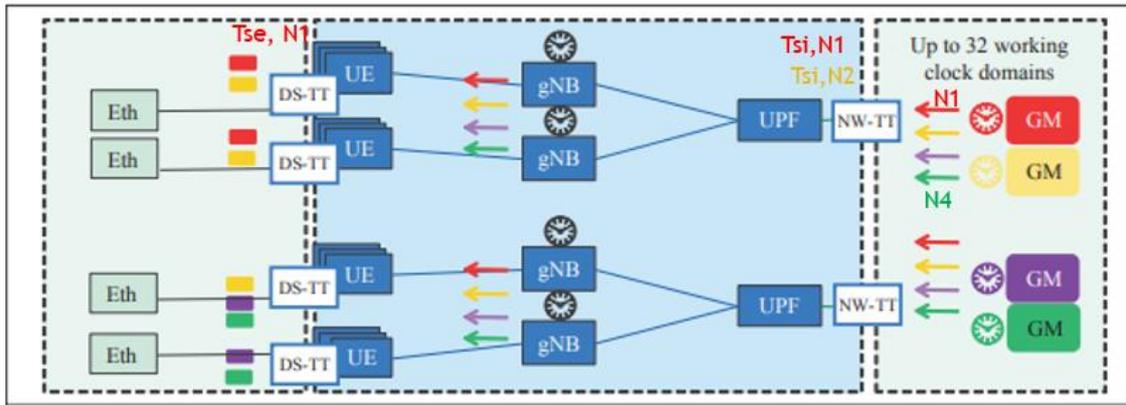


Figura 26: Sincronización temporal de la arquitectura TSN-5G con varios dominios temporales

### 2.3.5 Alta fiabilidad en redes TSN

Las redes TSN se caracterizan por su alta fiabilidad por lo que es necesario mejorar el sistema 5G para alcanzar esos niveles de confiabilidad. Para ello, existen diversas técnicas en el plano de usuario como la transmisión de múltiples antenas, el uso de múltiples portadoras o la duplicación de paquetes a través de enlaces de radio independientes. En un sistema 5G donde solamente se transmite una sesión PDU se puede mejorar la confiabilidad si se incrementa el número de rutas de transporte entre la RAN y el UPF, para conseguir redundancia e independencia. También se pueden utilizar múltiples antenas, múltiples portadoras y la duplicación de datos en enlaces radio independientes entre el UE y la RAN. Estas mejoras se detallan en la Figura 27, donde el apartado b muestra posibles rutas redundantes a través de nodos RAN y nodos UPF independientes, y el apartado c amplía la fiabilidad al utilizar UE redundantes.

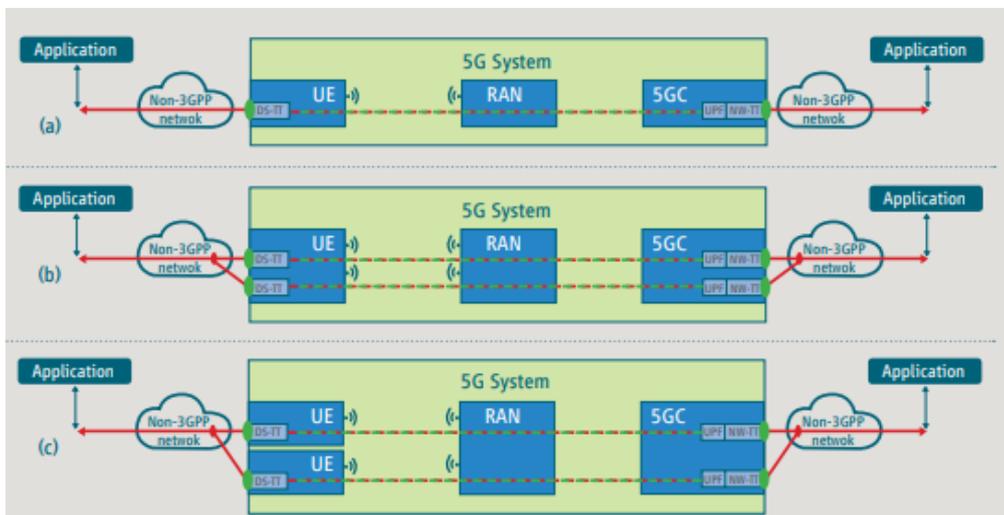


Figura 27: Posibles soportes para las comunicaciones ultra confiables [29]

Además, estas técnicas se pueden combinar con el uso de FRER para crear puentes redundantes. El sistema 5G puede crear puentes virtuales 5G independientes que al ser comunicados al CNC este configura el tráfico de flujo en las redes TSN para usar los puentes virtuales y los caminos redundantes.

También se puede introducir redundancia en el plano de control al utilizar una arquitectura basada en servicio (*Service Based Architecture, SBA*), que puede ser independiente de las soluciones de redundancia del plano de usuario.



## Capítulo 3

# TSN para la red de transporte de una red 5G

El diseño propuesto por 3GPP en el 3GPP *Release 16* y las mejoras introducidas por la alianza 5G ACIA para la integración de 5G en una red TSN, considerando el sistema 5G como un conjunto de conmutadores TSN virtuales, es un diseño válido, pero no permite definir la red de transporte como una red TSN. Por tanto, se deben definir nuevos escenarios que no están especificados en los estándares 3GPP para el desarrollo de una red de transporte TSN sobre una red 5G. Además, en la mayoría de las investigaciones y proyectos sobre la integración de TSN síncrono y 5G no se había contemplado la funcionalidad de la segmentación de red (*network slicing*), implementada en 5G, para las redes TSN. Esta funcionalidad permite que cada tipo de tráfico se pueda transmitir a través de una *slice* siempre que cumpla con las *QoS* requeridas y de este modo, el tráfico no se ve afectado por el resto de tráfico que circula por la red. Recientemente, se ha publicado en la revista *IEEE Access* un artículo [4] sobre este tema, es decir, se implementa una nueva arquitectura que permite interconectar un sistema de control y gestión de segmentos de red de la red móvil 5G con el plano de control de TSN. De esta manera, se habilita la tecnología TSN síncrono en la red de transporte para proporcionar las garantías de servicio de los diferentes tipos de tráfico. Para ello, se debe modificar un conjunto de propiedades necesarias para permitir la implementación de cortes de red en TSN. De esta forma, las redes TSN podrán controlar y administrar los diferentes segmentos de red extremo a extremo.

En este apartado se detalla el diseño y la arquitectura que se ha implementado en [4]. A partir de este diseño se considera que la segmentación de la red es una característica esencial para distinguir y separar los tipos de tráfico según sus calidades de servicio (QoS). Sin embargo, en [17] utilizan la tecnología de TSN síncrona para la red de transporte en lugar de TSN asíncrona. En este proyecto, se opta por el uso de TSN asíncrono ya que proporciona numerosos beneficios a la red. Algunas de las mejoras son la disminución de la complejidad de la red y el incremento de la escalabilidad y flexibilidad de la red. Por tanto, primeramente, se detallan las características de *network slicing* en las redes 5G, así como se produce la división por *slice* de los componentes físicos de la red. Y, por último, se explica el diseño de integrado utilizado basado en el descrito en [17] con *network slicing*.

### 3.1 Network Slicing en 5GS

Las redes 5G deben soportar simultáneamente el transporte de flujos procedentes de varios tipos de tráfico con características y requisitos de QoS heterogéneas y conflictivas entre sí. Para poder llevarlo a cabo se introdujo el concepto de *network slicing*. La segmentación de la red (*Network Slicing*) permite crear redes lógicas extremo a extremo que se adaptan a los diferentes casos de uso. Estas redes se implementan sobre unos mismos recursos físicos que son compartidos gracias a las tecnologías de SDN y NFV. Los diversos segmentos de red formados sobre una misma estructura están completamente aislados entre ellos, con control y gestión independientes. Además, se pueden crear nuevos segmentos según sea necesario, es decir, a demanda. Por ejemplo, en la Figura 28 se muestra una misma infraestructura física que implementa varios segmentos de red 5G adaptados y optimizados para distintos casos de uso como aplicaciones de usuario (*slice* azul), eSalud (*slice* verde) y comunicaciones máquina a máquina (*slice* roja).

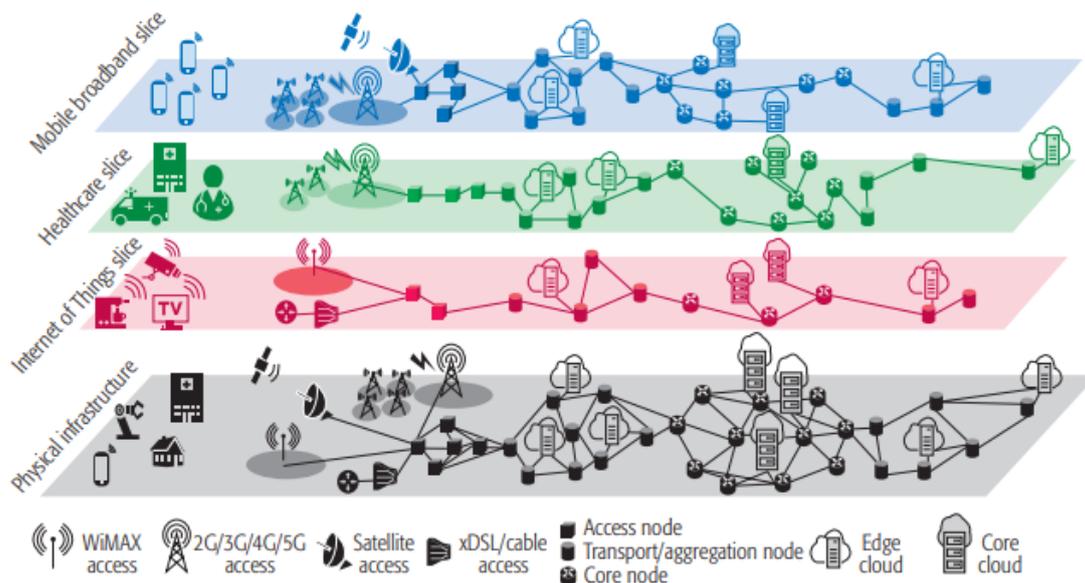


Figura 28: 5G network slices ejecutándose en una misma infraestructura física. [11]

Un segmento de red está formado por un conjunto de recursos que correctamente combinados crean una red lógica completa que cumple con los requisitos exigidos por los servicios del caso de uso soportado. Por lo tanto, un recurso es el conjunto de atributos o capacidades que pueden ser utilizadas para proporcionar un servicio. En las *network slices* existen dos tipos de recursos:

- **Funciones de red (*Network Function, NF*):** se trata de bloques funcionales que proporcionan las capacidades de red necesarias para efectuar y soportar el servicio demandado por el caso de uso. Los NF están implementadas como instancias de software que se ejecutan sobre los recursos de infraestructuras, y pueden ser físicos o virtualizados. Los NF físicas son una combinación de hardware y software de un proveedor específico formando un dispositivo físico tradicional para dicho propósito concreto. Mientras que un NF virtualizado tiene desacoplado su software del hardware donde se ejecuta.
- **Recursos de infraestructura (*Infrastructure Resources, IR*):** es el hardware y el software necesarios para la conexión y el alojamiento de los NFs. Está conformado por una serie de componentes físicos como capacidad de almacenamiento, recursos de

red, hardware de computación y puntos fijos para el acceso radio. Estos recursos y atributos deben dividirse lógicamente para poder implementar la segmentación de red. Para ello, se emplean mecanismos de virtualización que dividen un recurso físico en varios recursos virtuales para ser utilizados de igual forma que los físicos.

Según la terminología empleada por 3GPP, un segmento de red (*network slice*) es denominado como una instancia de segmento de red (*Network Slices Instance*, NSI), donde un NSI extremo a extremo está compuesto por varias instancias de subred (*Sub-network Instances*, NSSIs). Un NSSI es un conjunto de NF y sus correspondientes recursos son ordenados y configurados para conformar una red lógica. Por tanto, la red lógica completa puede estar formada por diferentes NSSIs. Además, un mismo NSSI puede ser compartido entre una o varios NSIs. Por ejemplo, un NSI extremo a extremo está formado por una RAN-NSSI que se conecta a un 5G-Core NSSI a través de una red de transporte NSSI (*Transport Network NSSI*, TN-NSSI).

La segmentación de red 5G debe desarrollar principalmente dos operaciones. La primera sería el control del tiempo de vida de las NSI, ya que es determinante para el control y la gestión de las acciones de aprovechamiento, instanciación y configuración, activación, desactivación, terminación y modificación de los segmentos. Mientras que la segunda es el proceso de asociación de un UE a un *slice* determinada.

La virtualización es un aspecto clave en la segmentación de la red, ya que permite descomponer los recursos en atributos siguiendo un criterio de selección predefinido, ocultando aquellos aspectos que no son importantes para la gestión y el uso de los recursos de manera útil. Estos recursos pueden ser físicos o estar previamente virtualizados. Si los recursos están ya virtualizados entonces se puede crear múltiples redes virtualizadas aisladas e independientes entre ellas y de la red física subyacente. Con la introducción de la virtualización se han creado nuevos actores de red y roles comerciales, agrupándose en tres tipos:

- **Proveedores de infraestructura (*Infrastructure Provider*, InP):** proporciona y gestiona la red física dada y sus recursos. Estos recursos son *data centers* (DCs) y/o WANs que se virtualizan para ser utilizados por uno o múltiples tenants a través de interfaces de programación.
- **Tenant:** alquila una red virtualizada formada por los recursos virtualizados proporcionados por uno o más InPs. Los tenants pueden ejecutar, administrar y proveer servicios de red a sus usuarios. Estos servicios de red están constituidos por una disposición de NFs, formada por NF individuales y un mecanismo que las conecta entre ellas.
- **Usuario final (*End User*):** consume los servicios o parte de ellos ofrecidos por los tenants, sin poder proporcionárselos a otros actores comerciales.

En la Figura 29 se observa el concepto de virtualización recursiva anteriormente mencionado, donde un tenant en una determinada capa actúa como InP en la capa inmediatamente superior. Esto ocasiona que un tenant ofrezca servicios de red a los usuarios finales y a otro tenant. Entonces el segundo tenant proporciona servicios de red más avanzados a sus usuarios finales.

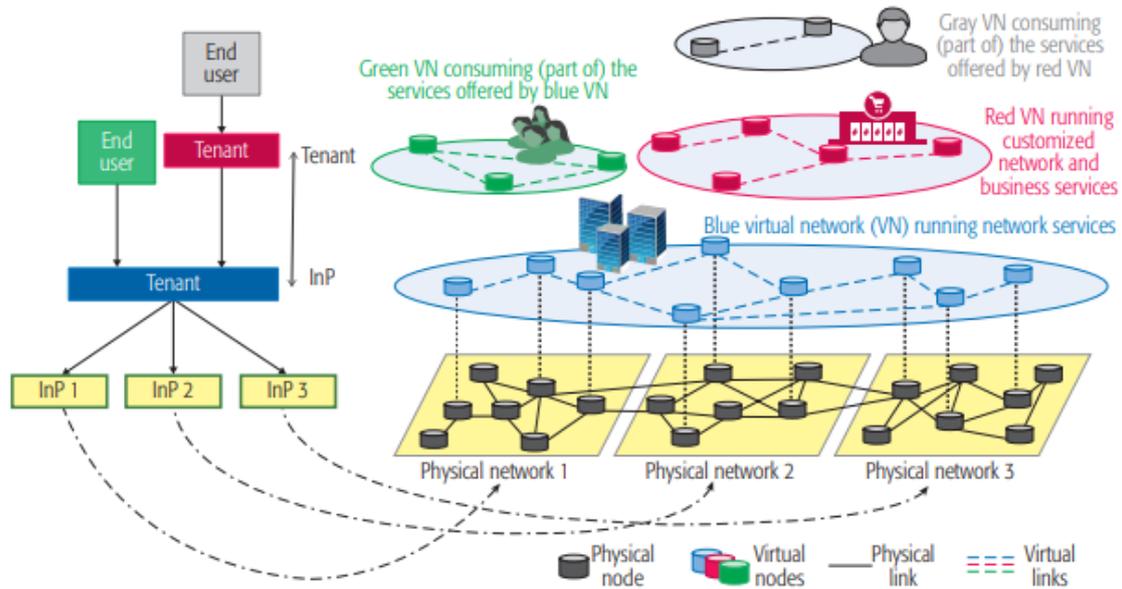


Figura 29: Principio de recursividad con InPs y tenants virtualizados. [11]

La orquestación de la red también es otro proceso importante, ya que la red al estar formada por componentes muy diversos es necesario un orquestador que coordine todos los procesos de la red para permitir la creación, gestión y entrega de servicios. Por tanto, el 3GPP ha definido en [12] una serie de entidades para la orquestación de los NSIs y NSSIs:

- **Función de gestión de servicios de comunicación (*Communication Service Management Function, CSMF*):** se encarga de mapear los requisitos de los servicios de comunicación a los requisitos de los segmentos de red.
- **Función de gestión de segmento de red (*Network Slice Management Function, NSMF*):** entidad que se encarga de la gestión y orquestación extremo a extremo de los NSI.
- **Función de gestión de subsegmentos de red (*Network Slice Subnet Management Function, NSSMF*):** se encarga de la gestión y orquestación de los NSSIs. Por ejemplo, en la NSI extremo a extremo formada por una RAN-NSSI, un 5G-Core NSSI y TN-NSSI, el RAN-NSSMF se encarga de la gestión y orquestas de RAN-NSSI. Mientras que el Core-NSSMF del 5G-Core NSSI y el TN-NSSMF del TN-NSSI.

La arquitectura y la comunicación entre interacción entre estas entidades se representa en la Figura 30. En ella se observa como en el plano de gestión se introduce la arquitectura ETSI MANO, ya que también puede ser incorporada. Esta arquitectura ETSI MANO permite el control de los recursos y servicios de red virtuales y físicos a través de la virtualización de la función de red (*Network Function Virtualized, NFV*). Sin embargo, las funciones de la RAN, el CU y el DU, se asumen que están virtualizadas mientras que la función RU es la única que no se virtualiza, ya que no se conoce cómo realizarlo y es un asunto abierto que está siendo investigado por OpenRAN y *Telecom Infra Project (TIP)*. Mientras que en la red de transporte la arquitectura ETSI MANO no orquesta los recursos de transporte necesarios, por lo que se centra en el control SDN.

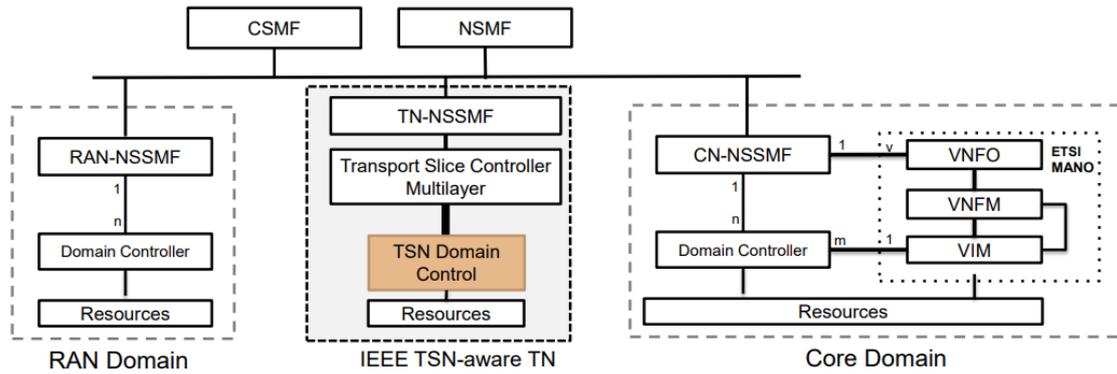


Figura 30: Orquestación y gestión de los slices de red. [4]

El objetivo de este proyecto es crear y configurar TN-NSSI basada en la red TSN asíncrona. Por lo que se profundiza en el funcionamiento de la red de transporte (*Transport Network, TN*) 5G, que interconecta los NFs desplegados en la RAN y en el *Core* de la red. Estos NFs también pueden estar desarrollados directamente en el hardware o ser VNFs ejecutadas en un sitio en la nube.

Con la virtualización de la RAN y la desagregación de la RAN y el Core junto con la división funcional las NFs RAN se reparten entre las nuevas componentes de la red. Es decir, las entidades CU, DU y RU alojan los NFs RAN mientras que los Core NFs residen en la red *Core* (5GC). Los enlaces TN que interconectan estas NFs son:

- **Fronthaul:** enlace que conecta la unidad radio (RU) y la unidad distribuida (DU), como se explicó anteriormente.
- **Middlehaul:** enlace que interconecta la unidad distribuida (DU) y la unidad centralizada (CU).
- **Backhaul:** enlace que conecta el CU con el núcleo de la red 5G (5GC).

## 3.2 Diseño

Las redes 5G presentan altas capacidades para cumplir con los requisitos exigidos por la gran variedad de escenarios de la industria 4.0. Sin embargo, no pueden garantizar los requisitos deterministas de latencia y el jitter extremo a extremo sin usar tecnologías como TSN o DetNet en la red de transporte para proporcionar conectividad entre sus componentes. El grupo de trabajo IETF DetNet está investigando nuevos mecanismos para proporcionar QoS determinística hasta capa 3, que engloba desde rutas explícitas, replicación y eliminación de paquetes hasta la protección frente a la congestión con sincronización extremo a extremo. No obstante, y aunque los estándares DetNet son independientes de TSN, suelen basarse en TSN para ofrecer QoS determinista hasta capa 2.

Numerosas son las investigaciones llevadas a cabo para encontrar el modelo adecuado de integración. El 3GPP propone un diseño de la posible arquitectura de integración entre 5G y TSN, pero parte de la premisa que la industria dispone originalmente de una red TSN en lugar de una red 5G. Este modelo, explicado en la Sección 2.3 *Integración de TSN y 5G*, considera el 5GS como un conjunto de puentes TSN. De esta forma, los estándares TSN pueden desarrollarse independientemente sin estar acoplados a la estandarización de 5G por el 3GPP. Por otro lado, se

emplea el protocolo TSN como se define en el 3GPP, pero también permite la integración con otros marcos tecnológicos como el soporte de TSN para MPLS, investigado por el IETF. Sin embargo, tampoco especifican la forma de implementación de dicho modelo.

Recientemente, en [4] se ha desarrollado un diseño donde se produce la integración entre la red 5G con TSN síncrono. Esta arquitectura conecta los sistemas de gestión de *network slicing* con el plano de control TSN síncrono basado en SDN. Este diseño mapea las instancias de segmentos de red definidas por 3GPP a la red de transporte TSN teniendo en cuenta los atributos de rendimiento de la red de transporte deseada. Para ello, presenta el controlador de segmento de red de transporte multicapa (*Multilayer Transport Network Slice Controller*, MTNSC) que interacciona tanto con el plano de control TSN y con la entidad de gestión de las *slices* de transporte. Además, un orquestador TSN se encarga de crear los segmentos de red TSN y de la gestión de su ciclo de vida.

En la mayoría de las investigaciones desarrolladas sobre la integración entre TSN y 5G para la Industria 4.0 emplea una red de transporte basada en TSN síncrono en lugar de asíncrono. Esto se debe a la sincronización temporal, ya que reduce considerablemente la latencia. Sin embargo, según los escenarios típicos de la industria 4.0 es preferible el uso de TSN asíncrono porque funcionan mejor con los tráfico deterministas aperiódicos (o esporádicos) y de mejor esfuerzo (*Best Effort*). Dado que TSN síncrono es adecuado para transmitir tráfico sensible al rendimiento con un patrón periódico como el tráfico generado por un sistema de control de bucle cerrado (*closed-loop*). Por tanto, se ha desarrollado un diseño de integración de 5G y TSN asíncrono que permite *network slicing*, basado en la arquitectura de [17]. Es decir, se ha creado una arquitectura donde la red que transporta los flujos de datos 5G deterministas entre la RAN y el CN utiliza TSN asíncrono.

En muchos escenarios de la Industria 4.0 es frecuente no conocer el número exacto de flujos que se deben asignar y sus características. Por tanto, esto supone que el problema de optimización asociado para configurar las redes TSN es estocástico en su naturaleza. Las entradas de este problema de optimización pueden ser una matriz de tráfico con el volumen de tráfico entre todos los pares de entrada y salida posibles y sus características, por ejemplo, la GFBR por cada 5QI. Los datos para crear la matriz de tráfico se pueden obtener de los datos históricos y del análisis predictivo. Existen dos posibles soluciones para resolver el problema de asignación de flujos en las redes TSN asíncronas, optando por un método fuera de línea para resolver el problema:

- **Métodos en línea:** la asignación de los flujos se realiza justamente después de su llegada a la red. Por lo que se puede emplear un algoritmo de optimización que encuentre la correcta asignación de cada flujo entrante. Estos métodos ofrecen una mayor flexibilidad que los fuera de línea, ya que los flujos de la misma clase o con los mismo 5QI pueden ser asignados por distintas configuraciones. Además, se adaptan con mayor agilidad a las condiciones cambiantes de la red.
- **Métodos fuera de línea:** resuelven el problema calculando una configuración de toda la red a largo plazo para cada uno de los tipos de tráfico. Para ello, primeramente, los flujos se agrupan en clase siguiendo un determinado criterio, como el parámetro 5QI o el 5QI y su *slice* en el TN 5G. Seguidamente se calcula la configuración de asignación para cada una de las clases. Estos métodos al presentar la misma configuración para todos los flujos de igual clase necesitan menos información de estado y el mecanismo de control de admisión solamente comprueba si se dispone de

los recursos suficientes, capacidad del enlace y espacio en el buffer, para los flujos de entrada.

La infraestructura de orquestación para las TNs basadas en ATS depende de los modelos de referencia ETSI ZSM (*Zero-touch network*) [25] y IETF ACTN (*Abstraction and Control of Traffic Engineered Networks*) [24]. Esta arquitectura permite que los clientes puedan crear y gestionar sus propias redes virtuales (TN *slicing*) abstrayendo la complejidad de la infraestructura física subyacente. También permite la coordinación de los diferentes dominios creados con el objetivo de garantizar la cohesión y la satisfacción de las configuraciones aplicadas en los distintos dominios. Los escenarios de la Industria 4.0 presentan requisitos muy exigentes con retardos de una cifra por lo que la topología empleada es reducida. Por tanto, se puede emplear una arquitectura totalmente centralizada de tipo SDN sin tener ningún problema de escalabilidad.

El modelo de referencia que define ACTN está formado por tres niveles para conseguir crear las redes virtuales como servicio en las TN. Por tanto, los componentes, que se muestra en la Figura 31, son:

- **Gestor de aprovisionamiento de tráfico (*Traffic Provisioning Manager, TPM*):** es el usuario de los servicios VN cuya función principal es transmitir todas las peticiones o comandos necesarios para la creación y la gestión de las VNs (TN *slices*). Esta función puede ser parte del plano de control de 5GC. También se encarga de calcular el volumen de tráfico previsto entre todos los pares de fuente y destino en las VNs. Por último, mapea los flujos 5G en las clases de tráfico TSN según las *slices* 3GPP y los 5QI definidos para cada flujo con el objetivo de que el nivel de servicio acordado sea concedido por la TN. La clase de tráfico TSN a la que ha sido asignada el flujo 5G se incluye en un campo inmutable de la cabecera del paquete.
- **Coordinador de servicio multidominio (*Multi-Domain Service Coordinator, MDSC*):** se encarga de organizar los diferentes dominios de TN. La TN puede presentar un solo dominio o tener múltiples dominios definidos, como por ejemplo la tecnología del proveedor o las zonas administrativas. También traduce las peticiones y los comandos de la TPM en un conjunto de parámetros específicos de la infraestructura de TN subyacente. Este conjunto de parámetros es utilizado por el controlador de TSN para configurar la red. Finalmente, separa los recursos de TN subyacente para ocultar la complejidad de configurar la red TPM.
- **Controlador TSN de redes definidas por software (SDN) (*Software-Defined Networking (SDN)/ TSN Controller, SDN/TSN-C*):** se encarga de del control de los dispositivos que conforman la red y de la supervisión y recolección de los datos de telemetría que se producen en la red. Además, debe determinar las configuraciones factibles de la red y asignar la cantidad suficiente de recursos a las redes virtuales para que puedan cumplir con las restricciones de rendimiento exigidos por los servicios 5G. Concretamente, el TSN-C para poder configurar correctamente los diferentes ATS que conforma una VN debe resolver el problema de asignación de flujos comentado anteriormente. Para solucionarlo se pueden implementar técnicas de análisis de datos e inteligencia artificial (IA). Por tanto, se puede considerar que actúa como un controlador de la red de aprovisionamiento.

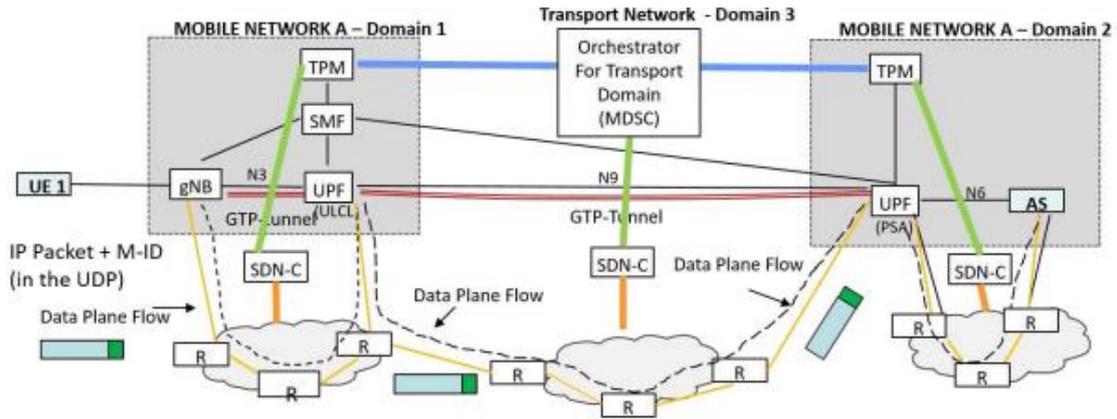


Figura 31: Arquitectura de la red de transporte backhaul. [33]

En la Figura 31 se observa que el segmento de transporte está dividido en tres secciones:

- **N3:** es el segmento o el dominio que conecta el gNB con el UPF.
- **N9:** es el segmento que conecta la UPF de una red inalámbrica y la UPF de otra red que se transmite a través de la red *backhaul*.
- **N6:** es el segmento entre el UPF y el servidor de aplicación (*Application Server, AS*).

Si además de los componentes descritos en el modelo de referencia ACTN [24] se tiene en cuenta el modelo de ETSI ZSM [25] se crea la infraestructura de gestión y orquestación mostrada en la Figura 32.

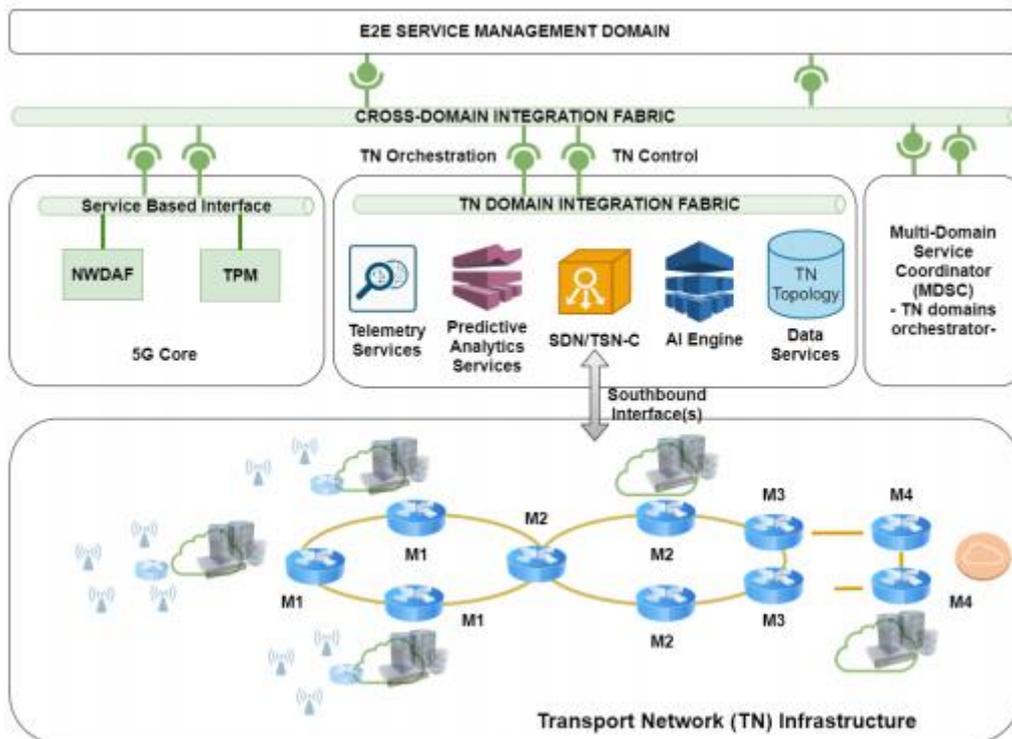


Figura 32: Arquitectura de gestión y orquestación de la red de transporte. [17]



## Capítulo 4

# Evaluación Experimental

Tras el diseño de una red de transporte TSN asíncrono en una red 5G con *network slicing* se debe resolver el problema de asignación de los flujos de las *slices* para cumplir con los requisitos de QoS exigidos. En este caso, se ha considerado implementar un optimizador que proporciona la distribución de retardo, la priorización de las *slices* en cada salto y la capacidad para cada *slice* óptimas. Este optimizador implementado es el descrito en [18] y pretende reducir la probabilidad global de rechazo de los flujos de la red. Para comprobar el correcto funcionamiento de las *slices* en el simulador con el optimizador óptimo se propone una serie de experimentos. Estos experimentos tratan de estudiar la granularidad para dar servicios a los tipos de tráfico contemplados en la Industria 4.0 según el número de *slices* adecuadas. Para ello, se estudia cómo influye a la probabilidad de rechazo global de la red el incremento del número de *slices* de un mismo tipo y la combinación existente entre las distintas *slices*. Se espera que al incrementar el número de *slices* se incremente la probabilidad de rechazo de flujo gradualmente y se vea afectada por los requisitos de retardo de las *slices*. Es decir, aquellas *slices* con mayores requisitos de retardo presentarán una peor probabilidad de rechazo que aquellas con menores requisitos.

Por tanto, en este apartado primeramente se especifica el escenario de simulación considerado y las suposiciones que se han considerado para su desarrollo. Seguidamente, se explica el simulador y las tecnologías utilizadas para resolver el problema de optimización. A continuación, se han definido las características de los tipos de tráfico más habituales en la Industria 4.0. Finalmente, se detalla los experimentos desarrollados y los resultados obtenidos.

### **4.1 Marco de evaluación**

En este proyecto se ha considerado un escenario industrial que presenta un área geográfica limitada y está implementada por una red 5G que da servicio a la industria, que puede estar o no integrada a la vez con una red TSN industrial. Esta red 5G utiliza la tecnología TSN asíncrono para implementar el *backhaul* (segmento que interconecta el UPF con la RAN). Esta red TSN asíncrono despliega una serie de *slices* para poder transmitir el tráfico de la red industrial, que presenta una serie de flujos con unas determinadas características de tráfico. Estos flujos nacen y mueren a través de un proceso de Poisson y con una llegada a cada uno de los nodos de la red TSN con una distribución exponencial. Para la gestión de la red de transporte se desarrolla un control de admisión según se ha explicado en las secciones anteriores.

Este diseño está basado en el descrito en [18] donde considera que una red TSN asíncrona completamente centralizada que está constituida por  $V$  puentes ATS que se interconectan a través de  $E$  enlaces simples. Se puede modelar directamente con un gráfico  $G = (\mathcal{V}, \mathcal{E})$ , donde  $\mathcal{V}$  y  $\mathcal{E}$  son los vértices (puentes) y las líneas (enlaces) del gráfico, respectivamente. El peso  $C^e$  de cada línea  $e \in \mathcal{E}$  representa la capacidad del enlace. Se considera que  $A \subseteq \mathcal{V}$  es el conjunto de puentes de acceso, es decir, los conmutadores TSN que conectan directamente con las fuentes y los destinos. Además, se ha considerado que  $P_{s,d}$  es el conjunto de caminos que interconecta cada par fuente  $s \in A$  y destino  $d \in A$ , donde un camino  $p \in P_{s,d}$  está comprendido por una secuencia de enlaces conectados por un puente TSN.

Cada puente TSN  $v$  presenta un puerto de entrada  $I^v$  y un puerto de salida  $O^v$ . En cada uno de los puertos de salida hay un ATS para manejar la transmisión de las tramas a través de un enlace. Cada ATS  $e$  tienen una cola de conformación  $S^e$  y una cola de prioridad  $P^e$ . Se ha definido  $Q$  tipos de clases de tráfico que cada una de ellas conforma una *slice*. Sus QoS se han definido en la Sección 4.3 *Escenario de evaluación*. Los flujos que pertenecen a las distintas *slices* está regulado y limitado por la función de la Sección 2.2.2 *TSN asíncrono*. Se ha supuesto que todos los flujos de una misma *slice* con tipo de clase  $q$  presentan el mismo tamaño de ráfaga  $b_q$ , mientras que la tasa promedio sostenible para los flujos de clase  $q$  se define como  $r_q$ . También, se asume que todos los flujos de una misma *slice* con clase  $q$  tienen la misma duración  $\tau_q$ . El retardo máximo que puede experimentar un flujo  $f$  cuando atraviesa los ATS se deriva de [5]. Por tanto, el retardo en el peor caso E2E  $D_f$  para transmitir cualquier flujo  $f$  en un camino  $P_{s,d} \in \mathcal{E}$  desde un origen  $s$  hasta un destino  $d$  se calcula con la fórmula:

$$D_f \leq \sum_{e \in \mathcal{E}_f} d_{f,e,P^e}^{max} = \sum_{e \in P_{s,d}} \left( \frac{b_{e,P^e}^{\wedge H} + b_{e,P^e}^{\wedge S} + l_{e,P^e}^{\wedge L}}{C_e - r_{e,P^e}^{\wedge H}} + \frac{l_f}{C_e} \right)$$

Donde  $b_{e,P}^{\wedge H} = \sum_{f \in F_{e,P}^H} b_f$  y  $r_{e,P^e}^{\wedge H} = \sum_{f \in F_{e,P^e}^H} r_f$  que son el burstiness agregado y la tasa de datos generados por un conjunto de flujos  $F_H$  con un nivel de prioridad mayor que el flujo  $f$  en el enlace  $e \in P$ ,  $b_{e,P^e}^{\wedge S} = \sum_{f \in F_{e,P^e}^S} b_f$  es el burstiness para el conjunto de flujos  $F_{e,P^e}^S$  con el mismo nivel de prioridad que el flujo  $f$  para el enlace  $e \in P$ ,  $l_{e,P^e}^{\wedge L}$  es el tamaño máximo de paquete de un conjunto de flujos con un nivel de prioridad inferior que el flujo  $f$ ,  $l_f$  es el tamaño de paquete máximo para los flujos  $f$  y  $C_e$  es la capacidad de transmisión de un ATS  $e$ .

En la arquitectura de red TSN completamente centralizada existe un controlador TSN lógicamente centralizado que se encarga del control y la supervisión de los diferentes puentes TSN que conforman la red. Por tanto, es el encargado de ejecutar el mecanismo de control de admisión, detallado en las secciones anteriores, para determinar que flujo de cada *slice* es aceptado o rechazado. Además, es el responsable de la gestión de toda la red de transporte, teniendo que configurar correctamente los ATS.

El optimizador implementado se describe en [18] cuyo objetivo es reducir la probabilidad de rechazo de los flujos que se transmiten en la red. Para ello, existe una serie de limitaciones:

- Se debe conocer el retardo extremo a extremo  $D_q$  de cada *slice* de clase  $q$ .
- La demanda de tráfico de cada uno de los enlaces  $e \in \mathcal{E}$  no puede superar su capacidad.
- No se debe superar el número de *buffers* conformados disponible para cada ATS para cumplir con las reglas QAR1, QAR2 y QAR3.

El problema de optimización también se resuelve en el mismo estudio [18] donde se aplican una serie de transformaciones para poder transformarlo en un programa no lineal de enteros mixtos convexo. Para ello, se definen previamente una serie de variables de decisión:

- $M_q^p$  indica el número promedio de flujos para una *slice* con una clase  $q$  que puede ser asignado a un camino  $p$  predefinido que interconecta una fuente y un destino dados. Esta variable multiplicada por la tasa promedio sostenible ( $M_q^p * r_q$ ) indica la capacidad reservada E2E de una *slice* de clase  $q$  en un camino  $p$ .
- $N_q^e$  es el número promedio de flujos de *slice*  $q$  que pueden ser atendido simultáneamente por un ATS  $e \in \mathcal{E}$ . La capacidad reservada para un *slice* de clase  $q$  para un ATS  $e \in \mathcal{E}$  determinado se calcula como  $N_q^e * r_q$ .
- $N_{q,p}^e$  es el número promedio de flujos de una *slice*  $q$  que puede ser atendido simultáneamente por un ATS con prioridad  $p$ .
- $U_{q,p}^e$  es una variable binaria que determina si el flujo con 5QI  $q$  es asignado a el nivel de prioridad  $p$ .
- $D_p^e$  indica el retardo en el peor caso experimentado por cualquier flujo asignado a un nivel de prioridad  $p$  en un ATS  $e \in \mathcal{E}$ .

El objetivo del optimizador es minimizar la probabilidad de rechazo de flujo global en la red, que se puede formar como:

$$\text{minimizar } \left\{ \sum_{r \in \mathcal{P}} \sum_{q \in \mathcal{Q}} \lambda_q^r * \alpha_q * \left( \frac{\lambda_q^r}{M_q^r * \mu_q + \lambda_q^r} \right) \right\}$$

Donde  $\left( \frac{\lambda_q^r}{M_q^r * \mu_q + \lambda_q^r} \right)$  indica el límite superior de probabilidad de rechazo para un sistema de colas M/G/c/c que modela la capacidad de cada camino  $r \in \mathcal{P}$  para asignar flujos de *slice*  $q$ . El parámetro  $\lambda_q^r$  representa la tasa de llegada media del camino  $r$  mientras que  $\mu_q$  indica la inversa del tiempo de vida medio  $\tau_q$  de los flujos de *slice*  $q$ . Este límite superior es ponderado por su respectiva  $\lambda_q^r$  para tener en cuenta la importancia de la probabilidad de rechazo de flujos en un camino  $r$  y una clase  $q$  en la probabilidad de rechazo global de la red.

Además, se formula una serie de restricciones para poder cumplir con las limitaciones anteriormente especificadas. Estas restricciones se especifican también en [18]. La siguiente restricción asegura que la capacidad del enlace  $e$  no sea superada por la tasa agregada sostenible del número de flujos asignados:

$$\sum_{q \in Q} N_q^e * r_q \leq C^e \quad \forall e \in \mathcal{E}$$

Por otro lado, se impone un límite superior a las variables  $N_{q,p}^e$  y  $N_q^e$ :

$$0 \leq N_q^e \leq \sum_{p=1}^P N_{q,p}^e \quad \forall e \in \mathcal{E}, q \in Q \quad 0 \leq N_{q,p}^e \leq N_{U,q}^e \quad \forall e \in \mathcal{E}, q \in Q, p \in [1, P^e]$$

La próxima función establece una relación entre las variables  $N_q^e$  y  $M_q^P$  con el objetivo de que la formulación del problema sea coherente.

$$N_q^e \leq \sum_{r \in P^e} M_q^r \quad \forall e \in \mathcal{E}$$

Para convertir el problema en un problema no lineal de enteros mixtos convexo y pueda ser resuelto por el un solucionador de optimización convexa se modifica algunas restricciones para eliminar los productos entre variables de decisión, dando como resultado las siguientes funciones de restricción. Las funciones que se indican a continuación fuerzan a que cada clase de *slice*  $q$  sea asignada a un solo nivel de prioridad  $p$  para un cada ATS  $e$ :

$$\sum_{p=1}^P U_{q,p}^e = 1 \quad 0 \leq N_{q,p}^e \leq U_{q,p}^e * N_{U,q}^e \quad \forall e \in \mathcal{E}, q \in Q, p \in [1, P^e]$$

$$0 \leq N_{q,k} \leq (1 - U_{q,p}) * N_{U,q}^e \quad \forall e \in \mathcal{E}, q \in Q, p \in [1, P^e] : k \neq p$$

Para garantizar que se cumple todos los requisitos de retardo E2E de cada tipo de *slice* se formula las restricciones posteriores, donde se incluyen nuevas variables de decisión  $z_p^e, w_p^e$  y otra variable de decisión auxiliar binaria  $y_{q,p}^e = U_{q,p}^e * D_p^e$ :

$$z_p^e = C^e - \sum_{q \in Q} \sum_{k=1}^{p-1} N_{q,k}^e * r_q \quad \forall e \in \mathcal{E}, p \in [1, P^e] \quad w_p^e \geq D^L * z_p^e \quad \forall e \in \mathcal{E}, p \in [1, P^e]$$

$$w_p^e \geq D^U * z_p^e + D_p^e * C^e - D^U * C^e \quad \forall e \in \mathcal{E}, p \in [1, P^e]$$

$$w_p^e \leq D^U * z_p^e \quad \forall e \in \mathcal{E}, p \in [1, P^e]$$

$$w_p^e \geq D_p^e * C^e + D^L * z_p^e - D^L * C^e \quad \forall e \in \mathcal{E}, p \in [1, P^e]$$

$$\sum_{e \in P} \sum_{p=1}^{P^e} y_{q,p}^e \leq D_q \quad \forall P \subseteq \mathcal{E}, q \in Q$$

$$D_p^e - (1 - U_{q,p}^e) * D^U \leq y_{q,p}^e \leq D_p^e + (1 - U_{q,p}^e) * D^U \quad \forall e \in \mathcal{E}, q \in Q, p \in [1, P^e]$$

$$0 \leq y_{q,p}^e \leq U_{q,p}^e * D^U \quad \forall q \in Q, p \in [1, P^e]$$

Finalmente, se introduce otra serie de restricciones para asegurar que el número de colas conformadas demandadas por cada ATS  $e$  sea inferior o igual que el número de colas conformadas que tiene disponibles. Para ello, también se introduce las variables binarias:

$$\delta_{q,k,p}^e = U_{q,p}^e * U_{k,p}^e \quad \forall e \in \mathcal{E}, q, k \in Q, q \neq k, p \in [1, P^e]$$

$$\bar{\delta}_{q,k,p}^e = U_{q,p}^e * (1 - U_{k,p}^e) \quad \forall e \in \mathcal{E}, q, k \in \mathcal{Q}, q \neq k, p \in [1, P^e]$$

$$\gamma_{q,k,p,m}^{e,i} = \delta_{q,k,p}^e * \bar{\delta}_{q,k,m}^i \quad \forall e \in \mathcal{E}, i \in I^e, q, k \in \mathcal{Q}, q \neq k, p \in [1, P^e], m \in [1, P^i]$$

Donde  $I^e$  es el conjunto de todos los saltos anteriores para el ATS  $e$ . Por tanto, las restricciones son:

$$\sum_{i \in I^e} \sum_{p=1}^{P^e} \sum_{q=1}^{|\mathcal{Q}|-1} \sum_{k=q+1}^{|\mathcal{Q}|} \sum_{m=1}^{P^i} \gamma_{q,k,p,m}^{e,i} + \sum_{p=1}^{P^e} \bigvee_{q \in \mathcal{Q}} U_{q,p}^e * I_q^e \leq S^e \quad \forall e \in \mathcal{E}$$

$$U_{q,p}^e - (1 - U_{k,p}^e) \leq \delta_{q,k,p}^e \leq U_{q,p}^e + (1 - U_{k,p}^e), \delta_{q,k,p}^e \leq U_{k,p}^e \quad \forall e \in \mathcal{E}, q, k \in \mathcal{Q}, q \neq k, p \in [1, P^e]$$

$$U_{q,p}^e - U_{k,p}^e \leq \bar{\delta}_{q,k,p}^e \leq U_{q,p}^e + U_{k,p}^e, \bar{\delta}_{q,k,p}^e \leq (1 - U_{k,p}^e) \quad \forall e \in \mathcal{E}, q, k \in \mathcal{Q}, q \neq k, p \in [1, P^e]$$

$$\delta_{q,k,p}^e - (1 - \bar{\delta}_{q,k,m}^i) \leq \gamma_{q,k,p,m}^{e,i} \leq \delta_{q,k,p}^e + (1 - \bar{\delta}_{q,k,m}^i), \gamma_{q,k,p,m}^{e,i} \leq \bar{\delta}_{q,k,m}^i \quad \forall e \in \mathcal{E}, i \in I^e, q, k \in \mathcal{Q}, q \neq k, p \in [1, P^e], m \in [1, P^i]$$

Todas estas restricciones junto con el objetivo de minimizar la probabilidad de rechazo global se implementan en el programa de optimización para resolver el problema de asignación de flujos.

## 4.2 Herramienta de evaluación

La parte experimental de este proyecto se desarrolla principalmente a través de tres herramientas de evaluación que permiten comprobar el funcionamiento del diseño de red *backhaul* TSN asíncrono. Para ello, la principal herramienta es un simulador que implementa tanto la topología de la red TSN asíncrona como el optimizador para resolver el problema de asignación de *slices* o la generación de las diferentes *slices*. Este simulador para resolver el problema de optimización necesita de un software para la programación convexa disciplinada (CVX) y un solucionador comercial (MOSEK).

### 4.2.1 Simulador

El funcionamiento del diseño detallado en el Sección 3.2 *Diseño* se comprueba a través del simulador detallado en [17]. Este simulador está basado en una prueba de concepto (*proof-of-concept*, PoC) y consta de los siguientes cinco módulos:

- **Generador de red (*Network generator*):** implementa la topología de la red utilizada para el funcionamiento. En este caso, se ha generado las tres principales topologías que se utilizan en la industria 4.0, es decir, una topología en anillo, de estrella y en cadena.
- **Generador de procesos de llegada de flujos (*Flows arrival process generator*):** se encarga de simular el proceso de llegada de los flujos a los nodos conforme a un proceso de renovación. Se puede utilizar tres distribuciones de tiempo entre llegadas de los flujos: exponencial, Erlang-2 e hiperexponencial. Las distribuciones de Erlang-2 y la hiperexponencial pueden aproximar cualquier función de probabilidad con un coeficiente de variación inferior y superior a uno, respectivamente. Para la separación de los flujos en *slices* estos deben presentar unas características de servicios similares,

para ello se ha definido una tabla donde se incluyen las restricciones de QoS para cada una de las *slices* que presentan los siguientes parámetros:

- *Rate* ( $r_j$ ): es el tráfico promedio en bits por segundo que genera un flujo dentro de una *slice*.
  - *Burstiness* ( $b_j$ ): es el número máxima de bits que un flujo de tráfico dentro de una *slice* puede generar en un momento determinado.
  - *Delay* ( $d_{max}$ ): es el retardo máximo que puede sufrir los flujos de una *slice* en la red TSN extremo a extremo.
  - *Jitter* ( $j_{max}$ ): es la máxima variabilidad que puede sufrir el retardo. En nuestro caso, se ha considerado dicho valor sea igual que el retardo con el objetivo de evitar posibles problemas.
  - *Reliability* ( $r_{min}$ ): es la fiabilidad que presenta los flujos del servicio proporcionado para evitar posibles fallos, como la pérdida de los paquetes. Esta probabilidad se mide en tantos por ciento (%).
  - *Priority* ( $prio$ ): es la prioridad asignada para cada uno de las *slices*. En nuestro caso no se tiene en cuenta ya que el programa de optimización intenta acomodar el mayor número posible de flujos de las distintas *slices*.
  - *Income* ( $inc$ ): indica el beneficio que obtendría un operador por incluir el *slices* dentro de su red, ya que acomodar unos tipos de servicios proporciona mayores beneficios que otro. En nuestro caso, no se tiene en cuenta por tanto todos los valores son 1 para evitar problemas.
  - *Duration* ( $dur$ ): es la duración media de los flujos de una *slice*. Es decir, indica el tiempo medio que un usuario está conectando, generando tráfico a la red. Se mide en segundos.
  - *Length* ( $l_{max}$ ): es el tamaño máximo de los paquetes de cada flujo que se genera en una *slice*. En nuestros casos, si las características de la *slice* indicadas en [9] no presentan este parámetro se considera que el tamaño máximo es 1500 bytes. Es este valor porque es el máximo permitido por *Ethernet* siendo TSN una extensión y teniendo el máximo valor permitido. Se mide en bits.
  - *5QI*: es el identificador de QoS 5G que se asocia a cada uno de los flujos de una *slice* que se transmite por una red 5G. Este parámetro está determinado por las propiedades del flujo como se muestra en la Tabla 1.
- **Optimizador de red (*Network optimizer*):** este módulo se encarga de resolver el problema de optimización para encontrar la asignación óptima de las prioridades de los flujos de las *slices* en cada uno de los saltos, la distribución de retardo y la minimización de la probabilidad de rechazo de flujo según el ratio de aceptación de los flujos en la red TSN asíncrona. El simulador utilizado presenta diferentes tipos de soluciones:
    - *eBaseline*: resuelve el problema según las prioridades 3GPP.
    - *eMindelay*: optimizador basando en minimizar el retardo que sufren los flujos.

- *eNeptuno*: optimizador que combina varios métodos para obtener una configuración a largo plazo del *backhaul* TSN asíncrono, minimizando la probabilidad de rechazo. Este optimizador está detallado en [7].
- *eOptimal*: optimizador que resuelve el problema maximizando el número de flujos que conforman distintas *slices* en la red TSN-5G. Este optimizador óptimo intenta minimizar la probabilidad de rechazo global que se produce en la red BN TSN asíncrono. La implementación de este optimizador se detalla en [18].

En este caso, se ha utilizado el optimizador *eOptimal* para la resolución del problema de optimización.

- **Control de admisión de *slices* (*Slices admission control*):** este mecanismo determina si la *slice* tiene los recursos suficientes para admitir un nuevo flujo entrante o una nueva *slices*. Para ello, es necesario que se ejecute en la fase de operación de la red.

### **4.2.2 CVX**

En [18] se detalla la formulación formal empleada para definir el problema de asignación de flujos en una red TSN asíncrona. Este problema tras las transformaciones propuesta por los autores se convierte en un programa no lineal de enteros mixtos convexo. Por lo que se necesita un marco o una infraestructura que permita su implementación en Matlab.

CVX es el marco empleado para programar el problema convexo en Matlab. Esto provoca que Matlab se convierta en un lenguaje de modelado, que permite definir las restricciones y los objetivos del problema con el uso de la sintaxis estándar de Matlab. Para resolver el problema es necesario utilizar la versión 2.0 de CVX que incluye el soporte para programas convexos de enteros mixtos (*Mixed-Integer Disciplined Convex Program*, MIDCP). Estos programas normalmente obedecen las reglas de programación convexa, pero puede suceder que una o más variables estén restringidas a asumir valores enteros o binarios. La mayoría de los modelos no convexos no se pueden expresar como MIDCP, ya que MIDCP no permite su resolución.

CVX también puede resolver problemas de programación geométrica (*Geometric Programming*, GP) con el uso del modo de GP especial. Los programas GP son no convexos, pero se pueden convertir en convexos a través de una determinada transformación. Sin embargo, esta transformación no es necesario realizarla antes, ya que el propio CVX permite la construcción del programa GP no convexo y lo transforma automáticamente a una forma convexa resoluble. Por último, traduce los resultados obtenidos del problema GP convexo al problema original.

Por consiguiente, CVX no es una herramienta de propósito general para resolver problemas de optimización no lineal, ni para verificar si un modelo es convexo o no. Entonces es necesario comprobar previamente si el modelo se puede expresar como un MIDCP o un GP. Si resulta que no es ninguno entonces no se puede usar CVX.

Esta herramienta es gratuita para un uso académico y comercial, pero se tiene que combinar con uno de los solucionadores gratuitos que ofrece, SeDuMi o SDPT3, para resolver el problema. Sin embargo, se pueden utilizar otros solucionadores, como Gurobi o MOSEK, si se adquiere la licencia profesional de CVX. Para resolver los MIDCP no se pueden utilizar todos los solucionadores, ya que algunos no los soportan. Además, aquellos que permiten la programación de MIDCP no garantizan una solución exitosa en un tiempo razonable para todos los modelos.

En nuestro caso se ha empleado el solucionador MOSEK por lo que es necesario obtener una licencia CVX. Esta licencia se ha adquirido sin ningún cargo por ser su uso puramente académico. Estas licencias académicas solo se pueden usar con fines educativos y de investigación sin

permitir el uso comercial, y caducan después de un año, pero pueden renovarse. Para obtener la licencia es necesario indicar la dirección de correo electrónico académico, el nombre de usuario y hasta dos ID de host, así como la Universidad a la que pertenece el usuario.

Una vez obtenida la licencia se debe descargar el paquete con todas las bibliotecas de CVX según el sistema operativo utilizado. En este caso se ha descargado para Windows. Tras la descarga se debe descomprimir el fichero en un directorio vacío y ejecutar el comando *cvx\_setup* en la línea de comandos de Matlab desde el correspondiente directorio.

### **4.2.3 MOSEK**

El solucionador utilizado para el problema de asignación de flujos es MOSEK. MOSEK ApS es considerado uno de los softwares comerciales principales para resolver problema de optimización convexa no lineal. Esta compañía fue creada en 1997 y está dirigida por el CEO fundador Erling Andersen y un asesor técnico Yinyu Ye, profesor de Stanford. Este solucionador puede soportar la variedad más amplia de posibles modelos de CVX que el resto de los solucionadores.

MOSEK Optimization Suite es el paquete software para Matlab capaz de resolver los problemas de optimización de gran escalar incluyendo problemas lineales, cuadráticos convexos, cuadráticos cónicos, semidefinidos y convexos en general. Aquellas variables con restricciones enteras son admitidas para todas las clases de problemas, excepto para los problemas semidefinidos y convexos en general. El MOSEK Toolbox para Matlab proporciona acceso a los optimizadores MOSEK del entorno Matlab.

Al igual que sucede con CVX, este solucionador requiere de una licencia para poder ser utilizado. Presenta también dos tipos de licencia, una para los usuarios académicos y otra para los usuarios comerciales. En este caso, se obtiene una licencia académica, que es gratuita y tiene una duración de un año, pero es posible la renovación. Las licencias académicas solo pueden ser utilizadas para fines educativos o de investigación en instituciones académicas. Para poder obtener el archivo de licencia es necesario indicar la dirección de correo electrónico académico, el nombre del usuario y de la institución educativa.

Para la instalación de MOSEK se debe tener instalado el paquete de CVX. Tras la obtención de la licencia de MOSEK está debe guardarse en la ubicación en la carpeta *mosek* -> *mosek.lic* del paquete descomprimido CVX. De esta forma se utiliza la versión MOSEK implementada en CVX, pero es posible realizar una instalación completa del solucionador si se desea utilizar fuera de CVX o una versión diferente a la proporcionada por CVX. Una vez añadida la licencia, en Matlab se debe volver a ejecutar el comando *cvx\_setup* para que pueda reconocer la nueva licencia de MOSEK, de esta forma aparecerá MOSEK entre los solucionadores disponibles.

En este caso, como se desea utilizar MOSEK como solucionador predeterminado, por tanto, es necesario cambiar la configuración ya que no se selecciona automáticamente como solucionador por defecto. Para ello, es necesario escribir en la línea de comandos de Matlab los comandos:

*cvx\_solver mosek* -> para cambiar el solucionador activo a MOSEK, pero solo para la sesión actual

*cvx\_save\_prefs* -> para guardar la selección en el archivo de preferencias de CVX

De esta forma, el solucionador MOSEK se selecciona como solucionador activo por defecto cada vez que se inicie Matlab.

### 4.3 Escenario de evaluación

Se ha considerado tres topologías de red diferentes para comprobar la operación de la red BN TSN 5G. Estas topologías son las típicas que se implementan en los escenarios industrias [30], y estas son en estrella, en cadena y en anillo. En la Figura 33 se muestran estas topologías, además se indican los caminos predefinidos que siguen los flujos de las *slices* y que interconectan cada par fuente-destino. Se considera un total de 5 nodos para todas las topologías, donde en la de estrella la fuente de los flujos de cada una de las *slices* son el nodo N2 y N4, y los nodos destino N3 y N5. Mientras que en la topología en cadena los nodos fuente son el N1 y el N5, y los nodos destino el N3 y N4. Por último, en la topología en anillo las fuentes son los nodos N2 y N5, y los destinos son los nodos N3 y N4.

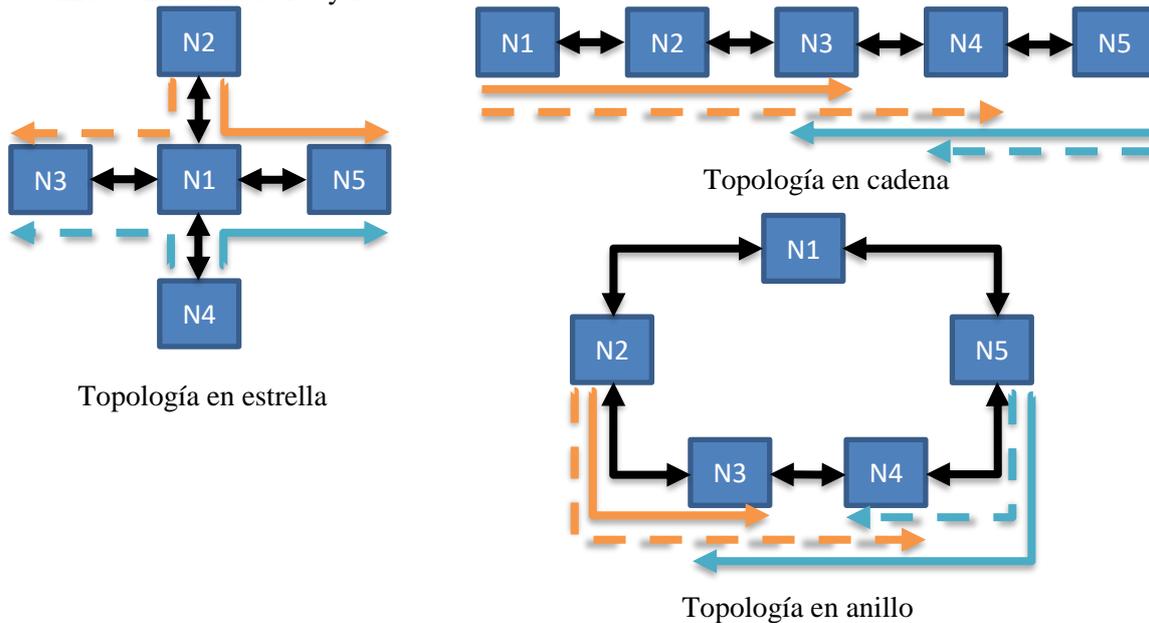


Figura 33: Topología de red industrial. [17]

Los enlaces que unen los nodos tienen una capacidad de un 1 Gbps y una distancia entre ellos de 200 m. Los flujos son transmitidos por el cable a la velocidad de la luz típica en estos medios, que suele ser en torno a  $2 * 10^8$  m/s. En los puertos de salida de cada uno de los nodos se implementa una instancia de ATS, donde cada uno de ellos incluye 100 buffers físicos para la conformación del tráfico, de modo que el número de estos no imponga restricciones en el número de niveles de prioridad implementables en el ATS. Además, se consideran 8 niveles de prioridad en los ATSs, ya que es el número máximo de niveles de prioridad por defecto que permite según se ha especificado en el estándar IEEE 802.1Q [26].

Para la configuración desarrollada se ha definido las propiedades, indicadas en el Sección 4.2.1 *Simulador*, de los diferentes *slices*, que se han obtenido de [27], [31] y [28]. Siguiendo el criterio de agrupación desarrollado en [27] los tipos de *slices* se agrupan según sus ámbitos de aplicación:

- **Automatización de fábricas (*Factory automation*):** se encarga del control automatizado, la supervisión y la optimización de los procesos, y los flujos de trabajo dentro de una fábrica. Los tipos de servicios que desarrolla suelen presentar los requisitos más altos de la infraestructura de comunicación subyacente, sobre todo en términos de disponibilidad, determinismo y latencia. Dentro de este ámbito se desarrollan diferentes tipos de servicios cuyo tráfico puede ser agrupado en *slices* independientes. Los servicios son:
  - **Control de movimiento (*Motion Control*):** Controla las partes móviles y/o giratorias de las máquinas, donde se envían una serie de mensajes que se deben

realizar de forma cíclica y determinista.

- **Comunicación control a control (*Control-to-Control Communication*):** Comunicación entre distintos controladores industriales. Normalmente, no existe una configuración fija, y los nodos de control que están en la red varían según el estado de las máquinas y la planta de fabricación, por tanto, el conexionado de los nodos de control es muy importante. Puede ser el tráfico de control procedente de los subsistemas de movimiento que se transmite entre controlados.
- **Robots móviles (*Mobile Robots*):** Uso de robot móviles con gran funcionalidad. Permiten una máxima flexibilidad en la movilidad, con cierta autonomía y capacidad de percepción (percibir y reaccionar entorno). Los robots móviles supervisan y controlan desde un sistema de control de guiado, que obtiene información del proceso para evitar colisiones, asignar tareas y gestionar tráfico.
- **Sustitución de enlaces cableados por inalámbricos (*Wired to Wireless link replacement*):** Se aumenta la flexibilidad en la configuración con enlaces inalámbricos. La comunicación entre dos máquinas que cooperan puede dividirse en tráfico periódico y aperiódico. Ambos están programados, y el tráfico por enlace varía según el número de máquinas que cooperen, pero el tráfico total en el equipo usuario no puede superar los valores preconfigurados.
- **Transporte cooperativo (*Cooperative carrying*):** Las piezas grandes y pesadas se transportan en robots móviles o AGV, que necesitan coordinarse entre ellos. Se alcanza con el control de los movimientos de los robots o AGVs. Esta comunicación requiere alta disponibilidad de servicio y una latencia ultra baja. El intercambio de comandos de control y retroalimentación de control se realiza con una comunicación determinista periódica con una red sensible al tiempo.

Las características de las *slices* que transporta cada uno de estos servicios se muestra en la Tabla 4.

Tipo de servicio industrial	Rate (r <sub>j</sub> ) [Kbps]	Burstiness (b <sub>j</sub> ) [bits]	Dmax (ms)	Jmax (ms)	Rmin (%)	Prio	Inc	Dur (s)	Lmax (bits)	5QI
Motion Control	1550e3	324*8	0.5	0.5	99.99	1	1	2880	324*8	1
Control-to-Control Communication	12.5e3	128*8	10	10	99.9	8	1	180	1000*8	8
Mobile robots	10e3	324*8	1	1	99.9	3	1	300	64*8	3
Wired to Wireless link replacement	50e3	448*8	0.85	0.85	99.9	2	1	300	1500*8	2
Cooperative carrying	2.5e3	180*8	2.5	2.5	99.9	5	1	3600	500*8	5

Tabla 4: Características de las slices de automatización de fábrica.

- **Automatización de procesos (*Process automation*):** se encarga del control de la producción y la manipulación de productos. Los tipos de servicios que desarrolla deben presentar un consumo de energía mínimo y gran seguridad para mejorar la eficiencia de los procesos de producción. De igual forma que el ámbito anterior, los servicios son:
  - **Control en bucle cerrado (*Closed-loop Control*):** en estos casos, la latencia y el determinismo son requisitos estrictos, así como la disponibilidad del servicio. El área de servicio es mayor que los casos de uso de control de movimiento y pueden no interactuar con la red pública. Pueden ser los sensores por planta que realizan mediciones continuas. Cuyos datos se transmiten a un controlador que toma la decisión de modificar el estado y/o características o no de los actuadores.
  - **Supervisión de procesos y activos (*Process and Asset Monitoring*):** es necesario numerosos sensores que permitan conocer las condiciones del proceso y del entorno, el estado de los activos y el inventario de materiales. Los datos son transportados para su observación y/o para su registro y análisis en las bases de datos. Los sensores industriales suelen estar limitados en tamaño, complejidad y/o consumo de energía. Su uso puede ser de área amplia y puede interactuar con la red pública.
  - **Gestión de activos de la planta (*Plant asset Management*):** se encarga del mantenimiento de los activos como bombas, válvulas, calentadores para que la planta siga funcionando. Es necesario reconocer la degradación y el autodiagnóstico continuo de todos los componentes para apoyar y planificar el mantenimiento. Las actualizaciones remotas del software mejoran y adaptan los componentes a las condiciones cambiantes y a los avances tecnológicos. Por tanto, las operaciones tienen una amplia área de servicio y puede ser necesaria la interacción con la red pública.

Las características de las *slices* que transporta cada uno de estos servicios se muestra en

la siguiente Tabla 5.

Tipo de servicio industrial	Rate (r <sub>j</sub> ) [Kbps]	Burstiness (b <sub>j</sub> ) [bits]	D <sub>max</sub> (ms)	J <sub>max</sub> (ms)	R <sub>min</sub> (%)	Prio	Inc	Dur (s)	L <sub>max</sub> (bits)	5QI
Closed-loop Control	2e3	448*8	10	10	99.9	9	1	2880	64*8	9
Process and Asset monitoring	1e3	186*8	100	100	99.9	10	1	2880	64*8	10
Plant asset management	2e3	180*8	100	100	99.9	11	1	180	255*8	11

Tabla 5: Características de las slices de automatización de procesos.

- **Interfaces hombre-máquina (*Human-machine interfaces, HMI*):** engloba todo el tráfico procedente de los dispositivos que permiten la interacción entre las personas y las instalaciones de producción. Los servicios desarrollados son:
  - **Paneles de control móviles (*Mobile Control Panels*):** Permiten interacción entre las personas y la maquinaria, utilizados para configurar, supervisar, depurar, controlar y mantener máquinas, robot, grúas, etc. Actualmente por seguridad está conectado a la maquinaria, pero con un enlace inalámbrico ultra fiable y de baja latencia se podría conectar los paneles móviles a las funciones de seguridad, para conseguir una mayor facilidad de uso y reutilizar paneles para otras máquinas. Es decir, se trata de una comunicación bidireccional periódica para el control remoto, por ejemplo: robots de montaje.
  - **Paneles operativos móviles (*Mobile Operation Panels*):** Consiste en el manejo y la supervisión de las máquinas, los robots móviles o las unidades de producción a través de un panel de control móvil que proporciona una mayor flexibilidad y comodidad. Un único panel puede controlar más de un sistema de producción debido a la movilidad. Proporciona información para la configuración, el control de las máquinas y la supervisión. Los datos de monitorización son menos críticos, pero soporta funciones críticas para la seguridad, como las paradas de emergencias, donde necesitan requisitos de latencias ultra bajas y transmisiones fiables.
  - **Realidad aumentada (*Augmented Reality*):** permite la supervisión de procesos y flujos de producción, las instrucciones paso a paso para tareas específicas (por ejemplo, en lugares de trabajo de montaje manual), y soporte *ad hoc* desde un experto remoto (por ejemplo, para las tareas de mantenimiento o servicios). El tráfico que genera es bidireccional entre los dispositivos de realidad aumentada y un servidor de procesamiento de imágenes, por ejemplo.

Las características de las *slices* que transporta cada uno de estos servicios se muestra en la Tabla 6.

Tipo de servicio industrial	Rate (r <sub>j</sub> ) [Kbps]	Burstiness (b <sub>j</sub> ) [bits]	Dmax (ms)	Jmax (ms)	Rmin (%)	Prio	Inc	Dur (s)	Lmax (bits)	5QI
Mobile Control Panels	4e3	324*8	4	4	99.9	6	1	300	250*8	6
Mobile Operation Panels	16e3	180*8	1	1	99.999	4	1	2800	100*8	4
Augmented Reality	20e3	2500*8	10	10	99.9	7	1	3600	1500*8	7

Tabla 6: Características de las slices de HMI.

- **Supervisión y mantenimiento (*Monitoring and maintenance*):** se encarga de la monitorización de los procesos y/o activos que pueden producirse en la producción industrial. Este proceso de monitorización no presenta un impacto instantáneo en los procesos y por lo tanto, el proceso de adquisición de los datos no presenta grandes exigencias en la latencia. En este caso solo hay un tipo de servicio:
  - **Acceso y mantenimiento remoto (*Remote Access and Maintenance*):** Mensajes para realizar el acceso y el mantenimiento a distancia de dispositivos y entidades, que pueden estar distribuidos geográficamente y ser fijos o móviles. Estos dispositivos poseen software/firmware que deben actualizarse ocasionalmente y se puede realizar en paralelo al proceso de producción u otros servicios de comunicación, sin ocasionar un impacto negativo en los servicios de comunicación de producción.

Las características de para esta *slice* se muestra en la Tabla 7.

Tipo de servicio industrial	Rate (r <sub>j</sub> ) [Kbps]	Burstiness (b <sub>j</sub> ) [bits]	Dmax (ms)	Jmax (ms)	Rmin (%)	Prio	Inc	Dur (s)	Lmax (bits)	5QI
Remote Access and Maintenance	1e3	500*8	100	100	99.9	12	1	180	256*8	12

Tabla 7: Características de las slices del ámbito de supervisión y mantenimiento

Para comparar los distintos tipos de tráfico con el tráfico generado por servicios que no requieren una alta fiabilidad y bajo retardo se ha definido una *slice* que transporta tráfico de tipo eMBB. Las características de está *slice* se muestran en la Tabla 8.

Tipo de servicio industrial	Rate (r <sub>j</sub> ) [Kbps]	Burstiness (b <sub>j</sub> ) [bits]	Dmax (ms)	Jmax (ms)	Rmin (%)	Prio	Inc	Dur (s)	Lmax (bits)	5QI
eMBB	5e3	4062*8	100	100	95	13	1	1200	1500*8	13

Tabla 8: Características de las slices del ámbito eMBB

Se ha supuesto que la tasa de llegada de flujos de una *slice* es similar para todos los tipos de servicios industriales considerados. Para los cuatro experimentos realizados se ha evaluado diez puntos de carga de trabajo (es decir, tasa de llegada de flujos) por cada par fuente-destino. Estos puntos corresponden con la utilización de los enlaces que oscila desde un 20% hasta el 190%. Esto se refiere a que si se aceptan todos los flujos entrantes generados por cada par fuente-destino la utilización de la capacidad del enlace variaría entre el intervalo de volumen de trabajo mencionado anteriormente.

## 4.4 Resultados

Para comprobar el funcionamiento del diseño se ha desarrollado dos experimentos con el objetivo de evaluar la heterogeneidad de las *slices* y el número máximo de *slices* posibles en una red.

- **Experimento 1:** se evalúa cómo afecta a la probabilidad de rechazo el incremento del número de *slices* dentro de la red. En este caso se consideran dos situaciones: i) existen 2 *slices* en la red, ii) existen 10 *slices*. Todas las *slices* son homogéneas, es decir, los flujos que se transmiten por los distintos segmentos son el mismo tipo de tráfico. La carga de tráfico de cada una de las *slices* es la misma y se reparte según el número de *slices* que presenta la red. Por tanto, para la situación de 2 *slices* con una capacidad máxima de enlaces de 1 Gbps se considera que cada segmento tiene una capacidad máxima de la *slice* de 500 Mbps mientras que para 10 *slices* es 100 Mbps por *slices*. Se ha considerado también *slices* para transmitir flujos de tipo eMBB, que presenta requisitos de retardo más lasos. En la Tabla 9 se muestra la tasa agregada total de la red a la que se alcanza una probabilidad de rechazo del 10%, explicado en el apéndice A, para todos los tipos de *slices* y las topologías, cuyas características se indican en la Sección 4.3 *Escenario de evaluación*.

<b>Topología RING</b>		
<b>Tráfico</b>	<b>2 slices</b>	<b>10 slices</b>
Motion Control	260 Mbps	*
Control-to-Control Communication	950 Mbps	610 Mbps
Mobile robots	430 Mbps	220 Mbps
Wired to wireless link replacement	230 Mbps	*
Cooperation Carrying	230 Mbps	150 Mbps
Closed-loop Access Control	520 Mbps	440 Mbps
Process Monitoring	540 Mbps	490 Mbps
Plant Asset Monitoring	510 Mbps	440 Mbps
Mobile Control Panels	500 Mbps	360 Mbps
Mobile Operation Panels	100 Mbps	*
Augmented Reality	380 Mbps	110 Mbps
Remote Access and Maintenance	530 Mbps	490 Mbps
eMBB	490 Mbps	330 Mbps

<b>Topología DAISY CHAIN</b>		
<b>Tráfico</b>	<b>2 slices</b>	<b>10 slices</b>
Motion Control	210 Mbps	*
Control-to-Control Communication	430 Mbps	200 Mbps
Mobile robots	430 Mbps	220 Mbps
Wired to wireless link replacement	250 Mbps	*
Cooperation Carrying	230 Mbps	150 Mbps
Closed-loop Access Control	510 Mbps	440 Mbps
Process Monitoring	530 Mbps	420 Mbps
Plant Asset Monitoring	510 Mbps	430 Mbps
Mobile Control Panels	500 Mbps	360 Mbps
Mobile Operation Panels	150 Mbps	*
Augmented Reality	150 Mbps	110 Mbps
Remote Access and Maintenance	530 Mbps	490 Mbps
eMBB	490 Mbps	330 Mbps

<b>Topología STAR</b>		
<b>Tráfico</b>	<b>2 slices</b>	<b>10 slices</b>
Motion Control	240 Mbps	*
Control-to-Control Communication	410 Mbps	160 Mbps
Mobile robots	430 Mbps	220 Mbps
Wired to wireless link replacement	200 Mbps	*
Cooperation Carrying	230 Mbps	150 Mbps
Closed-loop Access Control	510 Mbps	440 Mbps
Process Monitoring	530 Mbps	480 Mbps
Plant Asset Monitoring	510 Mbps	440 Mbps
Mobile Control Panels	500 Mbps	370 Mbps
Mobile Operation Panels	150 Mbps	*
Augmented Reality	380 Mbps	110 Mbps
Remote Access and Maintenance	530 Mbps	490 Mbps
eMBB	490 Mbps	330 Mbps

Tabla 9: Tasa agregada global para una probabilidad de rechazo del 10% según la topología.

\* Indican que la probabilidad de rechazo global inicial es superior a 0.1

En la Figura 34 se muestra estos mismos resultados, pero para las diferentes combinaciones de topología y número de *slices* de tipo eMBB.

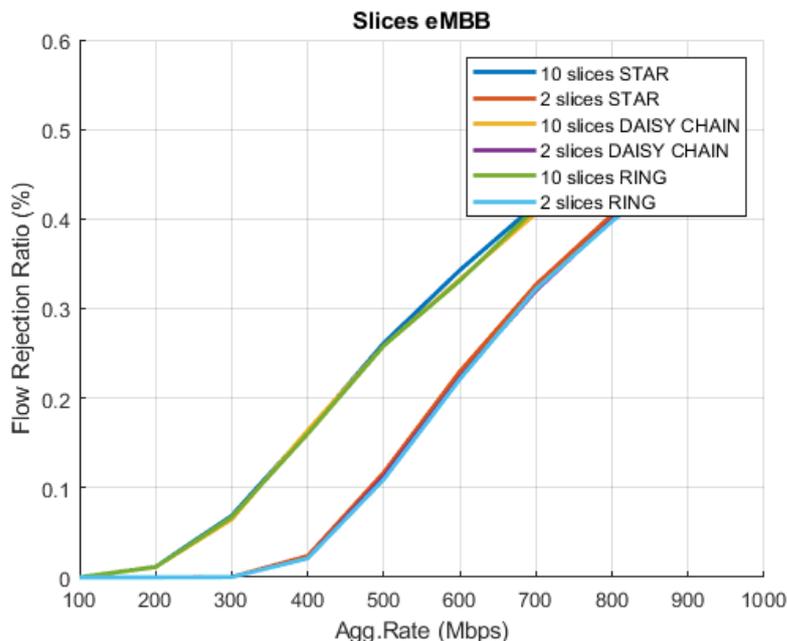


Figura 34: Representación de la tasa agregada en función de la probabilidad de rechazo para las *slices* de tipo eMBB según su topología y el número de *slices*.

Se observa que la tasa agregada para las distintas *slices* es independiente del tipo de topología empleada, exceptuando aquellos segmentos con requisitos de *delay* más estrictos, como *Motion Control*, *Wired to Wireless link replacement* o *Mobile Operation Panels*. En los casos donde se emplean principalmente estos tipos de segmentos se observa que se produce un incremento exponencial de la probabilidad de rechazo, llegando a alcanzar el 10% para 2 *slices* en valores muy bajos de la tasa agregada y se superan desde el inicio de la simulación para 10 *slices*. Esto se debe a las altas exigencias requeridas por los tipos de servicios en términos de retardo que provoca que la red no pueda cumplirlas por la congestión que se produce entre sus enlaces.

Por otro lado, se produce una reducción de la tasa agregada global para una probabilidad del 10% cuando se pasa de utilizar 2 *slices* a 10 *slices* en la red. Se debe principalmente a la multiplexación estadística, ya que con 2 segmentos de red cada *slice* tiene una mayor capacidad de los enlaces y, por tanto, una mayor ganancia por multiplexación a comparación con 10 *slices* que está más afectada por TDM.

Por tanto, para la Industria 4.0 no se debe implementar un número elevado de *slices* en la red, sobre todo en aquellos tipos de servicios que requieren altas restricciones de latencia y *jitter*.

- Experimento 2:** Este experimento es similar al anterior, pero en lugar de *slices* homogéneas se utilizan *slices* heterogéneas. Como se ha comprobado anteriormente es mejor emplear un menor número de *slices* para cumplir con los requisitos de retardo, por tanto, se ha considerado solo dos segmentos de red para cada tipo de servicio. Al ser los resultados independientes de la topología se considera un solo escenario con una topología en cadena con las mismas características que se ha explicado en la Sección 4.3 *Escenario de evaluación*. En este caso, también se comprueba como se ve afectada la probabilidad de rechazo de flujo global de la red cuando se combinan distintos tipos de *slices*. En la Tabla 10 se muestra la tasa agregada global a la que se obtiene la probabilidad de rechazo de 10% para la combinación de *slices* con QoS más estrictos y menos.

<b>Tráfico</b>	<b>2 slices-2 slices</b>
<b>Motion Control + Control-to-Control Communication</b>	350 Mbps
<b>Motion Control + eMBB</b>	400 Mbps
<b>Control-to-Control Communication + eMBB</b>	440 Mbps
<b>Mobile Robot +eMBB</b>	450 Mbps
<b>Mobile Robot +Control-to-Control Communication</b>	380 Mbps
<b>Motion Control + Mobile Robot</b>	420 Mbps

Tabla 10: Tasa agregada global para una probabilidad de rechazo del 10% según la combinación de *slices*

Se observa que para las *slices* que presentan unos requisitos de retardo más altos al combinarlas entre ella se reduce considerablemente la tasa agregada en la que se obtiene el 10% de la probabilidad de rechazo. Por tanto, se debe tener en cuenta la combinación de los tipos de *slices* a la hora de configurar la red.



## Capítulo 5

# Conclusiones y vías futuras

Tras la realización del proyecto, a modo de resumen, se presenta las conclusiones del trabajo realizado. Se ha llevado a cabo el estudio de las tecnologías de 5G y de TSN para su implantación en la Industria 4.0, haciendo gran hincapié en sus características. Así como los protocolos que implementa TSN para conseguir cumplir con los requisitos de comunicaciones con alta fiabilidad y bajo retardo. Según si se produce una sincronización temporal en los nodos de la red o no, existen dos versiones de TSN posibles: síncrono y asíncrono, donde se han analizado las principales diferencias entre ellas.

Además, se han estudiado las distintas investigaciones desarrolladas para la integración de ambas tecnologías y la inclusión de *network slicing* en la red de transporte basada en TSN. En este último caso se ha propuesto un diseño de integración de la red TSN asíncrona para soportar *network slicing*. Este diseño se basaría en la arquitectura desarrollada en [17] por los investigadores Prados-Garzón y Taleb, incluyendo la capacidad de segmentación de red.

Se ha desarrollado un simulador basado en el creado por [18] que implementa la separación en *slices* de la red. Para comprobar su correcto funcionamiento y evaluar aspectos de interés de *network slicing* en redes de transporte basadas en TSN asíncrono se han realizado dos experimentos, donde se estima la probabilidad de rechazo de flujo global de toda la red al incrementar el número de *slices* y la adaptación de la red a la acomodación de *slices* heterogéneas. Con estos experimentos se concluye que es mejor emplear un número reducido de *slices* teniendo en cuenta los requisitos de servicio para la industria 4.0.

Sin embargo, todavía existe una serie de limitaciones en el propio simulador. Por ejemplo, todavía no se contempla la posibilidad de dar más prioridad a los flujos de una determina *slices* para evitar ser rechazados. Es decir, en la situación actual el simulador rechaza los flujos de una *slice* independientemente del tipo de servicio que tenga asociado. Asimismo, con el optimizador empleado se proporciona la distribución adecuada de retardo, la priorización en cada salto y la capacidad en cada enlace para los distintos tipos de tráfico que se transmiten en las *slices*. Por tanto, no se puede determinar la asignación de los *slices* a los 8 PCP posibles contemplados por TSN asíncrono.

Por otro lado, este tema sigue teniendo grandes desafíos, ya que a pesar de que 3GPP *Release*

15 y 16 especifica un diseño de la integración no indica como se debe implementar. Por lo que existen varios temas que no se han estudiado:

- No sé conoce cómo se realiza la reserva de los recursos para los diferentes flujos con QoS distinta en el nodo gNB con el objetivo de reducir el retardo en la configuración temporal para el puente 5GS TSN. Al igual que la reserva de los recursos global en todo el sistema 5G para los flujos con QoS diferentes.
- La frecuencia con la que se debe enviar los mensajes *Sync Packet* en las redes TSN síncronas para evitar el retardo de los relojes GM en los diferentes nodos del sistema 5G.
- El mapeo entre las QoS de los tráfico 5G a las QoS de TSN.



## Capítulo 6

# Bibliografía

- [1] K. Campbell, L. Cruz, B. Flanagan, B. Morelli, B. O'Neil, S. Téral y J. Watson, «The 3G Economy "How 5G will contribute to the global economy",» IHS Markit, Noviembre 2019.
- [2] D. R. Prakash, «Transforming enterprise and industry with 5G private networks,» Qualcomm, 15 Octubre 2020.
- [3] P. M. Rost, D. Chandramouli y T. Kolding, «5G plug-and-produce How the 3GPP 5G System facilitates Industrial Ethernet deployments to fuel Industry 4.0 applications,» Nokia, 15 abril 2020.
- [4] S. Bhattacharjee, K. Katsalis, O. Arouk, R. Schmidt, T. Wang, X. An, T. Bauschert y N. Nikaiein, «Network Slicing for TSN-based Transport Network,» *IEEE Access*, pp. 62788-62809, 21 abril 2021.
- [5] J. Specht y S. Samii, «Urgency-based scheduler for time-sensitive switched ethernet networks,» de *2016 18th Euromicro Conference on Real-Time Systems (ECRTS)*, Julio 2016.
- [6] J.-Y. Le Boudec, «A theory of traffic regulators for deterministic networks with application to interleaved regulators,» *IEEE/ACM Transactions on Networking*, vol. 26, nº 6, pp. 2721-2733, Diciembre 2018.
- [7] J. Prados Garzón, T. Taleb y M. Bagaá, «Optimization of Flow Allocation in Asynchronous Deterministic 5G Transport Network by Leveraging Data Analytics,» *IEEE Transactions on Mobile Computing*, Julio 2021.
- [8] «IEEE Draft Standard for Local and metropolitan area networks—Bridges and Bridged Networks Amendment: Asynchronous Traffic Shaping,» *IEEE P802.1Qcr/D2.1*, pp. 1-152, Febrero 2020.
- [9] «TS 23.501 System architecture for the 5G System (5GS),» 3GPP, Diciembre 2017.
- [10] A. Larrañaga, M. C. Lucas-Estañ, I. Martínez, I. Val y J. Gozalvez, «Analysis of 5G-TSN Integration to Support Industry 4.0,» de *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Vienna, Austria, Sept. 2020.
- [11] J. Ordonez Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos Muñoz, J. Lorca y J. Folgueira, «Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges,» *IEEE Communications Magazine*, vol. 55, nº 5, pp. 80-87, Mayo 2017.
- [12] «TR 28.801 Telecommunication management; Study on management and orchestration of network slicing for next generation network,» 3GPP Technical report, Septiembre 2017.

- 
- [13] S. Rommer, P. Hedman, M. Olsson, L. Frid, S. Sultana y C. Mulligan, 5G Core Network Powering Digitalization, Academic Press, 2019.
- [14] Ericsson, «5G for business: a 2030 market compass. Setting a direction for 5G-powered B2B opportunities,» *Ericsson Technical Report*, Octubre 2019.
- [15] J. Farkas, B. Varga, G. Miklós y J. Sachs, «5G-TSN Integration for Industrial Automation,» *Ericsson Technology Review*, Julio 2019.
- [16] I. Gódor, M. Luvisotto, S. Ruffini, K. Wang, D. Patel, J. Sachs, O. Dobrijevic, D. P. Venmani, O. Le Mout, J. Costa-Requena, A. Poutanen, C. Marshall y J. Farkas, «A Look Inside 5G Standards to Support Time Synchronization for Smart Manufacturing,» *IEEE Communications Standards Magazine*, vol. 4, nº 3, pp. 14-21, Septiembre 2020.
- [17] J. Prados-Garzon y T. Taleb, «Asynchronous Time-Sensitive Networking for 5G Backhauling,» *IEEE Network*, vol. 35, nº 2, pp. 144-151, Marzo/Abril 2021.
- [18] J. Prados-Garzón, L. Chinchilla-Romero, P. Ameigeiras Gutiérrez, P. Muñoz Luengo y J. M. López Soler, «Asynchronous Time-Sensitive Networking for Industrial Networks,» *Institute of Electrical and Electronics Engineers (IEEE)*, 2021.
- [19] «How to build high performance 5G network with vRAN and O-RAN?,» *Qualcomm*, Febrero 2021.
- [20] A. Kaloylos, «A Survey and an Analysis of Network Slicing in 5G Networks,» *IEEE Communications Standards Magazine*, vol. 2, nº 1, pp. 60-65, 11 Abril 2018.
- [21] R. Ferrus, O. Sallent, J. Perez Romero y R. Agusti, «On 5G Radio Access Network Slicing: Radio Interface Protocol Features and Configuration,» *IEEE Communications Magazine*, vol. 56, nº 5, pp. 184-192, Enero 2018.
- [22] J. Prados-Garzón, T. Taleb y M. Bagaá, «Learnnet: Reinforcement learning based flow scheduling for asynchronous deterministic networks,» de *International Conference on Communications (ICC)*, Junio 2020.
- [23] A. Nasrallah, A. S. Thyagaturu, Z. Alharbi, C. Wang, X. Shao, M. Reisslein y H. ElBakoury, «Ultra-Low Latency (ULL) Networks: The IEEE TSN and IETF DetNet Standards and Related 5G ULL Research,» *IEEE Communications Surveys & Tutorials*, vol. 21, nº 1, pp. 88-145, 10 Septiembre 2018.
- [24] D. Ceccarelli y Y. Lee, «Framework for Abstraction and Control of TE Network (ACTN),» Internet Engineering Task Force (IETF), Agosto 2018.
- [25] C. Benzaid y T. Taleb, «Ai-driven zero touch network and service management in 5G and beyond: challenges and research directions,» *IEEE Network*, vol. 34, nº 2, pp. 186-194, Febrero 2020.
- [26] «IEEE Standard for Local and Metropolitan Area Network-Bridges and Bridged Networks,» IEEE Std 802.1Q-2018 (Revision of IEEE Std 802.1Q-2014), Julio 2018.
- [27] «TS 22.104 Service requirements for cyber-physical control applications in vertical domains,» 3GPP Technical specification, Diciembre 2018.
- [28] M. Granda Trigo, J. Ordonez Lucena, D. Camps, A. Garcia, G. Rigazzi, P. Ameigeiras, J. Prados Garzon, E. Aumayr, T. Cogalan, S. Yan, M. Ghorraishi y J. Guiterrez, «D2.1 Use Case Specifications and Requirements,» 5G-CLARITY Deliverable, 29 Febrero 2020.
- [29] 5G-ACIA, «Integration of 5G with Time-Sensitive Networking for industrial communications,» *White Paper 5GACIA*, Diciembre 2020.
- [30] 5G-ACIA, «Integration of industrial ethernet networks with 5g networks,» *White Paper 5GACIA*, Noviembre 2019.
- [31] 5G-ACIA, «A 5G traffic model for industrial use cases,» *White Paper 5GACIA*, Noviembre 2019.

- [32] Ericsson, «5G System: Enabling industry and society transformation,» Ericsson White Paper, Enero 2015.
- [33] Y. Lee, J. Kaippallimalil y R. Vilalta, «Extended ACTN Architecture to Enable End-To-End 5G Transport Service Assurance,» de *21st International Conference on Transparent Optical Networks (ICTON)*, Angers (Francia), Julio 2019.



## Apéndice A

# Cálculo de la probabilidad de rechazo de flujo al 10%

Una probabilidad de rechazo de flujo indica el porcentaje de flujos que se ha rechazado en la red debido a no ser capaz de cumplir con sus QoS. El 10% es el valor en el que la red alcanza el 0.1 de flujos rechazados. La tasa agregada total de la red a la que se obtiene dicho valor se calcula representando la probabilidad de rechazo obtenida del simulador de todas las *slices* según las configuraciones de red y *slices* considerados en función de la tasa agregada global de la red. Este último valor también se calcula en el simulador y depende de la utilización del enlace para cada una de las *slices* y de su capacidad. La tasa agregada para una determinada combinación de *slices* y configuración de red es el valor correspondiente en el eje X para una probabilidad de rechazo de 0.1 en el eje Y. En la Figura 35 se muestra el proceso detallado anteriormente para una topología en estrella con 2 *slices* de tipo *close-loop Access control* con una capacidad de enlace de 1 Gbps.

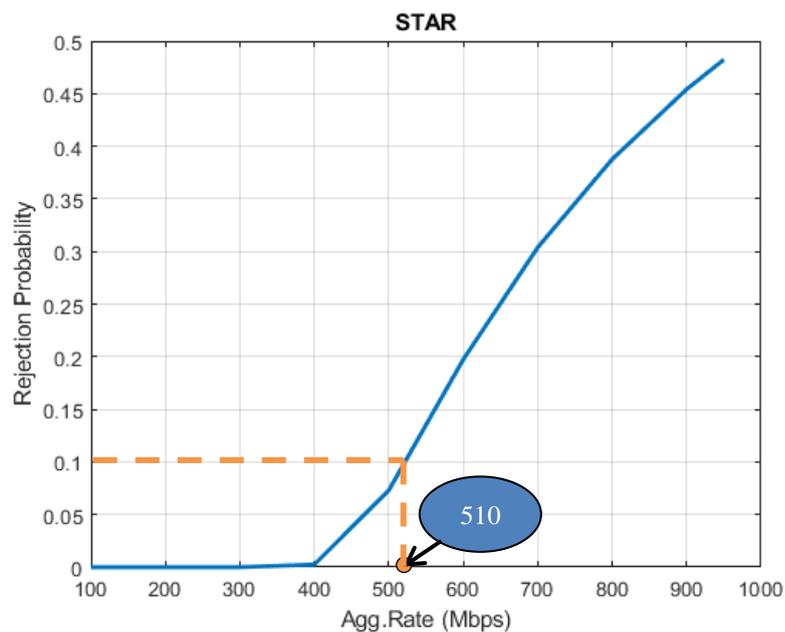


Figura 35: Tasa agregada (Mbps) en función de la probabilidad de rechazo del 10%

