SIMAGRO: Un prototipo para la detección de anomalías en entornos IoT para el sector agroalimentario

Marta Fuentes-García
COSCYBER
FIDESOL
Granada, Spain
mfuentes@fidesol.org

Roberto Magán-Carrión
NESG
Universidad de Granada
Granada, Spain
rmagan@ugr.es

Celia Fernández
COSCYBER
FIDESOL
Granada, Spain
cfernandez@fidesol.org

David Álvarez
COSCYBER
FIDESOL
Granada, Spain
dalvarez@fidesol.org

Marina Torres
COSCYBER
FIDESOL
Granada, Spain
mtorres@fidesol.org

***Resumen*—El sector primario es uno de los más relevantes en Andalucía. Una de las áreas más importantes dentro de este sector es la agricultura, destacando la producción de aceituna, frutas y hortalizas tropicales, además de los cultivos ecológicos (estos últimos suponen la mitad del total en España). Tras los años que se han sucedido de crisis, uno de los pilares fundamentales para que se reactive este sector es la optimización de las técnicas de cultivo, lo que implica la necesidad de una transformación digital profunda. Por esta razón, la sensorización de plantaciones agrarias y la implantación del IoT (del inglés, *Internet of Things*) como mecanismo de monitorización de los cultivos supone un gran avance para las entidades que lo están implantando.**

ÉGIDA es la primera Red de Excelencia Cervera para la privacidad y la seguridad de los datos. Uno de los objetivos de esta Red es concienciar sobre la necesidad de llevar a cabo una digitalización segura. En este sentido, existe una alta implicación con la securización activa de los entornos IoT, concretamente en el sector agroalimentario. En este contexto, y fruto de la colaboración activa entre la Universidad de Granada (UGR) y Fidesol, se ha llevado a cabo el desarrollo un prototipo para la detección de anomalías en entornos IoT para el sector agroalimentario. Este prototipo aplica por primera vez el sensor MSNM (MSNM-S) en un escenario IoT. El objetivo de este artículo es doble: por un lado, probar el funcionamiento de Atenea Lab y, por otro, presentar los resultados de la evaluación de este prototipo y resolver las siguientes cuestiones: *i)* ¿Es aplicable MSNM jerárquico a entornos IoT? y *ii)* ¿Cómo afecta la configuración de MSNM-S a entornos IoT? Además, se pretende identificar posibles puntos de mejora para continuar evolucionando tanto el prototipo obtenido para IoT como el sensor de MSNM.

***Index Terms*—IoT, IoT agroalimentario, detección de anomalías, laboratorio, transformación digital, transferencia de conocimiento**

Tipo de contribución: *Transferencia*

I. Introducción

Según distintas fuentes, el sector agrario se planteó como un sector estratégico para favorecer la recuperación económica en Andalucía tras la crisis derivada del COVID19 [1], [2]. Uno de los pilares fundamentales para que se produzca esta reactivación es la optimización de las técnicas de cultivo, lo que implica una necesidad de la digitalización del sector [2]. Son varios los trabajos que proponen el uso de IoT en el sector agroalimentario y que lo enmarcan dentro de la Industria 4.0 [3], [4]. De hecho, existen algunas soluciones que proporcionan un producto completo, desde la sensorización hasta la visualización de datos para optimizar la cosecha.

Algunos de los principales beneficios de la monitorización de plantaciones agrarias mediante dispositivos IoT son [5], [6]: optimización en la detección de alteraciones en el estado del cultivo, incremento de la productividad y correlación entre datos sensóricos. Sin embargo, el inconveniente es que, en ocasiones, la inversión que supone el diseño y la implantación de un sistema IoT no es fácilmente identificable como un potencial beneficio a medio-largo plazo. Además, es necesario disponer de expertos para dotar de valor los datos obtenidos, así como interpretarlos de manera adecuada. Por último, los sensores y dispositivos IoT desplegados en este tipo de contexto son más vulnerables, dado que suelen encontrarse en la intemperie y pueden sufrir no sólo las inclemencias climáticas, sino también accidentes que los dañen y alteren los datos.

I-A. Identificación de problemas de seguridad e implicaciones en entornos agrarios con ecosistemas IoT

Un problema importante cuando se afrontan las amenazas de seguridad de la información es que el tiempo necesario para comprometer un sistema o red es mínimo (del orden de segundos o minutos) si se compara con el tiempo que se tarda en detectar y reaccionar frente a un ataque (que puede llevar desde días hasta meses) [7]. Por eso, es esencial reducir el tiempo de detección y respuesta, así como es deseable que los mecanismos de detección también permitan la adecuada priorización de las alarmas.

Si bien es cierto que los datos que se envían mediante los sensores que componen un ecosistema IoT en el sector agroalimentario no son datos de carácter personal, sufrir un ataque podría tener numerosas consecuencias negativas para la empresa: desde el espionaje industrial hasta la alteración de datos y modelos. Teniendo en cuenta que este ecosistema IoT podría ser parte de un sistema ciber-físico y que los modelos podrían ser utilizados para regular aspectos como la cantidad de pesticida necesario para optimizar la producción, la gravedad de estos ataques es aún mayor, pues podrían llegar a poner en peligro la salud de las personas.

I-B. Principales amenazas y retos en entornos IoT

Algunos de los ataques más comunes son: fuerza bruta [8], [9], DoS (del inglés, *Denial of Service*) [10], [8], [9],

utilización del dispositivo atacado como plataforma de ataque hacia otros dispositivos [11], [8], inyección de información falsa [10], espionaje (captura de tráfico) [11], [9], suplantación de identidad [9], MiTM (del inglés, *Man in The Middle*) [11], [9], y obtención de datos [10], [8]. Otros ataques son los relacionados con la ocupación ilegal del canal de comunicación (*jamming*), el reenvío selectivo de paquetes, la redirección de paquetes o tráfico, la inundación con paquetes «hello», el análisis del tráfico de red, el acceso no autorizado (malicioso o no), los ataques de criptoanálisis [11], o la instalación de *malware* [11], [9]. En muchos casos, los atacantes se aprovechan de puertas traseras o vulnerabilidades y se realiza ejecución remota de código [10], [8]. Por otra parte, cuando existen *routers*, la mala configuración de estos puede conllevar el robo de información confidencial, el uso de la red para acciones ilegales, la vinculación con todo lo que ocurra en la red (tanto legal como ilegal), la infección de los dispositivos con *malware*, o la apropiación del ancho de banda [9].

En [12], [8] se detallan algunos casos reales de ataques relacionados con IoT. Por ejemplo, desde un frigorífico se puede enviar SPAM [12], pero también existen ataques a infraestructuras críticas que quedan comprometidas [12], [10], [8]. Tal es el caso de los ataques que tuvieron lugar en 2020 contra el sistema de aguas de Israel. El objetivo de los atacantes era alterar el nivel de clorina para producir el envenenamiento de la población [13], [14]. Más recientemente (febrero de 2021), un sistema de depuración de agua en Florida, fue también atacado, incrementando la cantidad de hidróxido de sodio (desinfectante) aceptable para el consumo humano. La intrusión fue detectada a tiempo y el agua no llegó a consumirse, pero pone de manifiesto la fragilidad y exposición de las infraestructuras críticas [15]. Finalmente, el sector agroalimentario también se encuentra en alerta por los crecientes ataques (especialmente de *ransomware*) en épocas clave para las cosechas [16]. Así, el mayor reto en seguridad IoT es evitar que los atacantes tomen el control del ecosistema [10].

Son muchos los protocolos de comunicación existentes tanto para IoT como para IIoT (del inglés, *Industrial IoT*). Aunque la mayoría de estos protocolos implementa mecanismos de seguridad, casi todos se centran en soluciones criptográficas para proporcionar autorización y autenticación. Por tanto, aún es necesario desarrollar nuevos protocolos y soluciones de seguridad que satisfagan los requisitos y retos identificados por distintos autores. De todos los trabajos estudiados se extrae como conclusión que el requisito de seguridad para el que existen menos soluciones es la disponibilidad, siendo dicho requisito de especial relevancia en IIoT (donde se enmarca el sector agroalimentario). Otro de los grandes desafíos consiste en proporcionar soluciones resilientes frente a ataques y que superen las restricciones de recursos propias de este tipo de redes.

Por supuesto, también es importante securizar el acceso a los sistemas y el almacenamiento de los datos. Por lo general, las soluciones existentes se basan en la aplicación de técnicas de cifrado y control de accesos.

Así mismo, y de cara a la defensa frente los anteriores ataques y amenazas a la seguridad, es de gran interés abordar la detección de ataques e intrusiones de forma similar a como se hace en sistemas TIC (Tecnologías de la Información y la Comunicación). Por tanto, el diseño de IDS (del inglés, *Intrusion Detection System*) y SIEM (del inglés, *Security Information and Event Management*) adaptados a IoT/IIoT es otro reto importante. En este sentido, la detección de anomalías es una tendencia cada vez más extendida, ya que permite detectar tanto ataques conocidos como nuevos (también conocidos por su nomenclatura en inglés, ataques de *zero-day*).

En este trabajo nos centramos en esta última parte, con el objetivo de estudiar y solucionar algunos de los posibles problemas de ciberseguridad derivados del uso de IoT en entornos agrarios. Para ello, partimos de un escenario simplificado, en el que se considera que la solución IoT desplegada utiliza protocolos abiertos y es configurable. En concreto, se despliegan e integran por primera vez varios sensores que implementan la metodología MSNM [17] para la detección de anomalías en el tráfico de red a través de la herramienta MSNM-S[1] [18] (desarrollada por la Universidad de Granada) en un escenario IoT construido utilizando el Laboratorio de IoT (Atenea Lab, desarrollado por Fidesol). Así, las principales aportaciones de este artículo son:

1. Identificación de las principales amenazas de seguridad en entornos IoT/IIoT.
2. Validación del funcionamiento de Atenea Lab con un prototipo de detección de anomalías en el tráfico de red para el sector agroalimentario.
3. Aplicación por primera vez de MSNM-S en un entorno realista IoT en el marco del acuerdo de trabajo entre Fidesol y la Universidad de Granada.
4. Validación de la viabilidad práctica de MSNM-S para la detección de ataques de DoS en ecosistemas IoT del sector agroalimentario.
5. Identificación de puntos de mejora y trabajo futuro sobre MSNM-S (con énfasis en entornos IoT).

El resto del documento se organiza como sigue: en la Sección 2 se describe brevemente el trabajo relacionado con esta propuesta, en concreto, se explican los dos recursos principales sobre los que se sustenta: el Atenea Lab y la metodología MSNM. En la Sección 3 se detalla el caso de uso (escenario IoT simplificado para el sector agroalimentario) y la configuración del escenario que permite llevar a cabo el prototipo en Atenea Lab. Se presenta la solución propuesta para adaptar, desplegar e integrar varios sensores MSNM en el escenario IoT propuesto. En la Sección 4 se muestra la experimentación llevada a cabo para validar el funcionamiento de la solución, así como el análisis y visualización de los resultados obtenidos. Finalmente, en la Sección 5 se exponen las conclusiones derivadas del trabajo realizado, así como el trabajo futuro relacionado con el mismo.

II. Materiales y métodos

II-A. Atenea Lab

Atenea Lab es un laboratorio virtual desarrollado por Fidesol para simular entornos IoT mediante software. Gracias a este laboratorio, se pueden simular escenarios variados, permitiendo, entre otras cosas, probar soluciones de ciberseguridad.

[1] https://github.com/nesg-ugr/msnm-sensor/

El Atenea Lab está diseñado y desarrollado de forma modular, de manera que es posible añadir funcionalidades en forma de "plugin".

Actualmente, Atenea Lab permite la simulación de dispositivos (incluyendo aspectos como la frecuencia de envío de datos y módulos para modelar el comportamiento), la comunicación mediante el protocolo MQTT[2] y el almacenamiento de los valores recibidos en una base de datos MongoDB[3]. **Atenea Lab tiene como objetivo representar un escenario realista** en el que se reproducen comportamientos similares a los que se encontrarían en un entorno operativo real. Se pretende así obtener una simulación de alto nivel, mediante la abstracción de algunas características que permitan obtener resultados realistas, ignorando aspectos como la marca o el modelo del dispositivo, y simulando de forma sencilla el comportamiento de éste. Cada dispositivo se compone de varios sensores, por lo que es posible configurar un amplio abanico de escenarios: desde dispositivos con un único sensor hasta combinaciones de dispositivos con distinto número de sensores y comportamientos. Se prevé que Atenea Lab esté disponible como entorno de experimentación y colaboración para proyectos IoT en el futuro.

II-B. Metodología MSNM

MSNM (del inglés, *Multivariate Statistical Network Monitoring*) [17] es una extensión de la metodología MSPC (del inglés, *Multivariate Statistical Process Control*) que permite la monitorización para la seguridad en redes de comunicaciones. MSNM se suele aplicar junto con técnicas basadas en variables latentes, como PCA (del inglés, *Principal Component Analysis*), utilizando un par de estadísticos complementarios que permiten la monitorización indirecta de un alto número de variables que representan al entorno monitorizado. Estos estadísticos se denominan **Q** (**Q-st**) y **D** (**D-st**), y se calculan a partir de la descomposición PCA de los datos de calibración para construir un modelo de operación normal (NOC, del inglés, *Normal Operation Condition*) [19]. Primero se identifican y resuelven las anomalías asignables y se genera el modelo NOC para, posteriormente, llevar a cabo la monitorización y poder así detectar las anomalías [20]. Con ayuda de estos estadísticos se pueden detectar patrones anómalos de comportamiento cuando sus valores exceden ciertos límites de control, como se puede comprobar en la Sección IV a través de los denominados gráficos de monitorización.

Tanto MSPC como MSNM se podrían enmarcar dentro del paradigma de aprendizaje automático no supervisado. MSNM permite combinar datos de tráfico con otras fuentes de seguridad [21] y ha demostrado un rendimiento de detección comparable con otras metodologías de aprendizaje automático del estado del arte [22]. La ventaja más relevante de MSNM en relación a estas metodologías es su interpretabilidad y su capacidad de ayudar en el proceso de diagnóstico [22], [23]. Al igual que otras metodologías de aprendizaje automático, y a diferencia de MSPC, MSNM necesita realizar pasos de *parsing* y fusión. Esto se debe a que los datos de red proceden de registros de distintos sensores y en distintos formatos, por lo que es necesario procesar y transformar los datos para que tengan un formato uniforme e interpretable. Así, MSNM consta de cuatro pasos: 1) *Parsing*, 2) Fusión, 3) Detección, 4) Diagnóstico.

II-B1. Variantes MSNM. MSNM jerárquico: Desde que MSNM se presentó en 2016, la metodología ha sido extendida, proponiendo mejoras en los pasos existentes, así como añadiendo otros nuevos [20]. MSNM jerárquico es una alternativa para la fusión de datos que se presentó por primera vez en [24]. La fusión jerárquica consiste en calcular los estadísticos en distintas capas o niveles de una jerarquía, en lugar de en un único nivel. Tanto los estadísticos como los datos se pueden combinar en distintos puntos de la jerarquía utilizando uno o más integradores. En general, la fusión jerárquica de los datos presenta los siguientes beneficios:

- Permite la priorización e identificación de la localización y/o la fuente de la anomalía.
- Reduce el tiempo consumido y el volumen de datos necesarios para llevar a cabo la monitorización.
- Incrementa la escalabilidad del sistema, ya que se pueden añadir varias fuentes a la arquitectura de la jerarquía, proporcionando distintos escenarios.
- Mejora la privacidad, ya que es posible aplicar fusión de alto nivel en las capas más altas de la jerarquía, evitando enviar las características o los datos en crudo al integrador.

Los ecosistemas IoT típicamente se organizan en distintos niveles (p.e. dispositivos, *edge*, nube) y disponen de recursos limitados. Por eso, los beneficios de MSNM jerárquico anteriormente mencionados hacen que sea de especial interés su uso en escenarios IoT.

II-B2. MSNM Sensor (MSNM-S): Para implementar la metodología MSNM, en la Universidad de Granada se desarrolló un prototipo que permite generar modelos estáticos y dinámicos a partir de datos de tráfico de red. Se trata del MSNM Sensor (MSNM-S) [18][4].

MSNM-S es altamente configurable, flexible y escalable y admite cualquier tipo de fuente de datos. En concreto, permite implementar tanto la variante de fusión estándar como la jerárquica de MSNM mediante la combinación de varios sensores. Además, MSNM-S hace uso del FCParser [5], una herramienta desarrollada para implementar la fase de *parsing* de la metodología MSNM y que sigue el enfoque FaaC (del inglés, *Feature as a Counter*). Esta herramienta transforma las variables originales, procedentes de fuentes de información en crudo (p.e. flujos de tráfico NetFlow) en otras derivadas. Las nuevas variables consisten en contadores de ciertos valores que contienen las variables originales. Esta transformación se lleva cada ciertos intervalos predeterminados de tiempo (por defecto, un minuto). Así, una variable derivada podría ser el número de paquetes contados en dicha ventana temporal, útil para detectar posibles ataques a la disponibilidad de un sistema. Con esta transformación, se homogeneiza la información obtenida a partir de fuentes heterogéneas. Esto permite su fusión y uso en modelos de aprendizaje automático, como PCA. De esta forma, los tres primeros pasos de la metodología MSNM quedan cubiertos.

[2]https://mqtt.org/

[3]https://www.mongodb.com/atlas/database

[4]https://github.com/nesg-ugr/msnm-sensor/

[5]https://github.com/josecamachop/FCParser

III. Detección de anomalías en entornos IoT. Caso de estudio: sector agroalimentario

Diseñar un sistema IoT completo no es una tarea sencilla. Para ello, es necesario tener en cuenta distintos factores, como el contexto de aplicación, las necesidades de monitorización y las condiciones de comunicación. Para el sector agroalimentario, partiremos de un caso de uso simplificado y simulado que, en adelante, llamaremos SIMAGRO (del inglés, *SIMulated AGROnomy*). Para definirlo, se han considerado los siguientes factores:

- Contexto de aplicación: sector agrario, sin especificar sub-sector. Suponemos una extensión de cultivo de entre una y varias hectáreas.
- Necesidades de monitorización: estado y condiciones del producto y entorno. Suponemos que es interesante obtener datos meteorológicos, así como del suelo de cultivo.
- Condiciones de comunicación: suponemos que no existen restricciones específicas en cuanto a comunicación. Podemos suponer que lo importante es recibirlos de forma robusta, aunque ello suponga la posibilidad de perder algunos paquetes y recibir otros duplicados.

Estas asunciones han permitido llevar a cabo un primer estudio sobre las necesidades de ciberseguridad en entornos agroalimentarios, independientemente del sub-sector concreto de aplicación (p.e. cultivo de hortalizas, frutas, o aceituna). En las siguientes sub-secciones se describen brevemente los elementos considerados para plantear este escenario y se justifica la elección de estos.

III-A. Sensores IoT

El escenario IoT para SIMAGRO se ha configurado según lo indicado en la Tabla I. Se puede observar que hay dos dispositivos: uno para el ambiente (D1) y otro para el suelo (D2), con un sensor de humedad y temperatura cada uno. Los valores de cada sensor tienen un rango definido y se simulan mediante variaciones incrementales por tramos horarios:de 7h a 15h, de 15h a 19h y de 19h a 7h. Se supone que estos sensores son básicos y tan sólo miden los datos indicados. Los datos son enviados haciendo uso del protocolo de comunicación seleccionado. Además del valor medido, también se envían: identificador de dispositivo y sello temporal (*timestamp*) de la medición.

III-B. Protocolos de comunicación

Para que los datos medidos por los sensores tengan sentido, deben ser enviados a un centro de cálculo que pueda procesarlos. De esta forma, se pueden obtener predicciones para, por ejemplo, optimizar la producción o detectar si la misma está evolucionando según lo previsto. En este caso, se asume que el centro de cálculo será un ordenador/servidor desplegado en las instalaciones físicas de la empresa SIMAGRO. Para este prototipo, se ha elegido MQTT, ya que su funcionamiento es simple y eficaz (publicación/suscripción), y es apto para un consumo reducido de ancho de banda y de recursos en general. MQTT funciona mediante una arquitectura cliente-servidor. En esta arquitectura de comunicación, los publicadores son los que envían los datos. En este caso, los publicadores son los sensores. Los suscriptores son los que reciben o leen la información. Para ello, se suscriben a los temas de interés (*topics*) en el servidor MQTT (*broker*). En este prototipo, sólo hay un suscriptor, que se encuentra en el ordenador que procesa los datos. Como se trata de un escenario simplificado, se ha asumido que no se dispone de dispositivo de borde (*edge*), y que el preprocesamiento, el tratamiento y el análisis de los datos se llevan a cabo en el servidor desplegado en las oficinas de SIMAGRO.

III-C. Almacenamiento

Ya en el servidor de las oficinas de SIMAGRO, los datos deben ser almacenados para su posterior tratamiento y uso. Para ello, se ha supuesto que se dispone de una base de datos no relacional. La razón es que estas bases de datos son más económicas y flexibles. Lo más normal es que se elijan bases de datos de tipo MongoDB[6] o Cassandra[7], dada su flexibilidad y capacidad de trabajo con cantidades masivas de datos. En SIMAGRO se ha optado por una base de datos MongoDB, pues tiene un carácter generalista que es útil no solo para SIMAGRO, sino para el propio Atenea Lab, donde se simula el entorno IoT. Además, dispone de otras características que la hacen atractiva para ambos proyectos, entre las que destacan: seguridad avanzada, concurrencia y alta disponibilidad y escalabilidad[8].

III-D. Adaptación, despliegue e integración de los sensores de MSNM en el Atenea Lab

El prototipo se ha desarrollado conectando cinco máquinas virtuales en el Atenea Lab: una para cada uno de los dispositivos IoT, otra para el *broker*, otra para el suscriptor (que lee los datos publicados por los dispositivos) y otra para la base de datos (que almacena los valores recibidos). La Figura 1 muestra la disposición de los dispositivos IoT y la base de datos de forma simplificada, así como la estructura de comunicación y los sensores de MSNM utilizados para la detección de anomalías en este prototipo. A la izquierda se representan los dispositivos, que contienen los dos sensores descritos anteriormente, en el centro se encuentran el *broker* MQTT y el suscriptor. A la derecha se representa la base de datos. Cada uno de estos elementos (a excepción del suscriptor) tiene desplegado un MSNM-S y un mecanismo de captura y exportación de flujos de tráfico NetFlow como principal fuente de información de cada sensor de MSNM-S.

Finalmente, todo el escenario se encuentra desplegado en un entorno virtualizado en el que se han asignado direcciones IP individuales dentro de la misma red a cada elemento. Mediante el acceso a la base de datos, se ha podido comprobar que Atenea Lab funciona correctamente, pues se reciben y almacenan los valores generados por los sensores IoT, que coinciden con lo definido durante la configuración del escenario (Tabla I).

IV. Experimentación

En esta sección, se valida el comportamiento de MSNM-S en un escenario IoT realista, gracias a la infraestructura

[6]https://www.mongodb.com/es

[7]https://cassandra.apache.org/_/index.html

[8]https://www.knowledgehut.com/blog/data-science/cassandra-vs-mongodb
https://www.openlogic.com/blog/cassandra-vs-mongodb
https://www.mongodb.com/compare/cassandra-vs-mongodb

Tabla I: Configuración del escenario IoT en SIMAGRO.

Dispositivo	Sensor	Cantidad	Id	Valores
Ambiental (D1)	temperatura ambiente	1	STA	$[17-35]°C$
	humedad ambiente	1	SHA	$[30-60]\,\%$
Suelo (D2)	temperatura suelo	1	STS	$[17-35]°C$
	humedad suelo	1	SHS	$[30-60]\,\%$

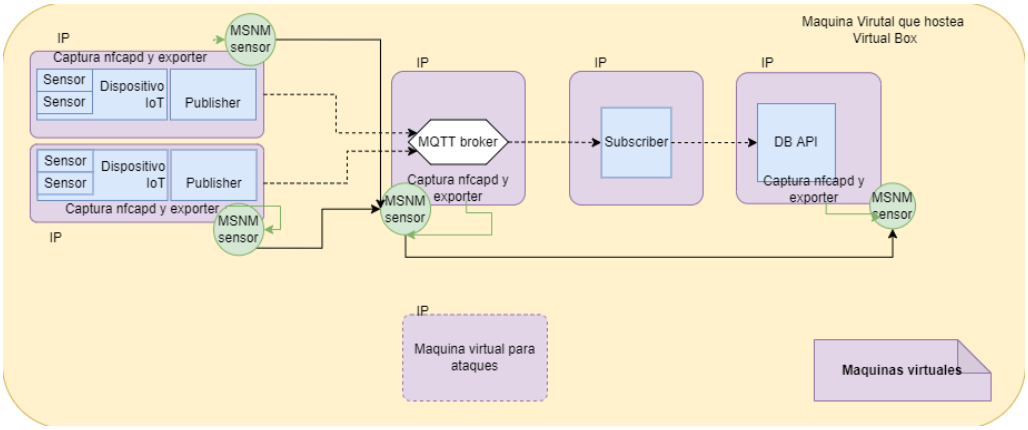


Figura 1: Disposición de los dispositivos, comunicación y almacenamiento de los datos enviados en el entorno IoT propuesto para SIMAGRO. Despliegue de los sensores de MSNM para la detección de anomalías.

proporcionada por el Atenea Lab. Para ello, se han diseñado dos experimentos (**E01** y **E02**) con el objetivo de validar la viabilidad práctica de la herramienta en entornos IoT frente ataques de disponibilidad del sistema.

El comportamiento de MSNM-S se configura mediante archivos de configuración YAML. Para cada sensor MSNM, se deben indicar parámetros tales como el origen de los datos a monitorizar o el destino donde se almacenarán las características extraídas. Para el origen de los datos también es necesario indicar si la fuente es local, es decir, el propio elemento donde se encuentra el sensor (en este caso, NetFlow) o remota, es decir, correspondiente a un sensor de MSNM externo (en este caso, los estadísticos de monitorización calculados en dicho sensor). En ambos casos hay que indicar la dirección IP de las fuentes. Por otra parte, también hay que indicar parámetros del funcionamiento del modelo PCA que incluye el sensor de MSNM, como el **número de componentes principales**, el **tipo de modelo** (estático o dinámico) o el **tipo de pre-procesamiento**, entre otros. La selección de estos parámetros requiere de un estudio previo del tipo de datos a monitorizar.

Para seleccionar el número de componentes principales, se ha llevado a cabo un análisis exploratorio de datos, incluyendo el estudio del número de componentes mínimo que captura la mayor parte de la varianza de los datos (p.e. el 80 % de la varianza de los datos). Para SIMAGRO, el número de componentes principales es 5 cuando la calibración es dinámica en los sensores MSNM (E01) y 2 para la calibración estática (E02).

Respecto al tipo de modelo, influyen factores como la variabilidad en los patrones del tráfico de red y la adaptabilidad que se desea del modelo generado. En la alternativa estática se parte de un conjunto de observaciones normales para calibrar el modelo, mientras que en la opción dinámica el modelo se va actualizando de manera periódica. En el caso del MSNM-S, se utiliza el método de actualización dinámica EWMA (del inglés, *Exponentially Weighted Moving Average*) [21]. Los principales parámetros de EWMA son: el factor de olvido (λ) y el número de observaciones con el que se re-calibra el modelo (B). λ toma valores entre 0 y 1, donde 1 representa la ausencia de olvido (se utilizan todas las observaciones pasadas). En este caso, se han probado distintas alternativas para estudiar cómo afecta cada tipo de modelado a los entornos IoT: todos los sensores MSNM se calibran dinámicamente (E01) y todos los sensores MSNM se calibran estáticamente (E02). El pre-procesamiento suele ser autoescalado, ya que los datos de tráfico de red son de naturaleza heterogénea.

En la Tabla II se resumen las características de los tres experimentos llevados a cabo para probar el funcionamiento del prototipo.

Tabla II: Configuración de MSNM-S para los dos experimentos

Id. experimento	#PCs	Calibración	F. olvido	#Obs. cal.
E01	5	Dinámica	$\lambda = 0{,}1$	$B = 30$
E02	2	Estática	-	-

Finalmente, otro aspecto importante que es posible configurar son las características extraídas (contadores) de las fuentes locales monitorizadas en cada sensor. Estas características pueden ser definidas manualmente, mediante conocimiento experto, o de forma automática, gracias a la aplicación de métodos de aprendizaje automático. Las características son procesadas por FCParser y se almacenan en un archivo de configuración en formato YAML que, posteriormente, es interpretado por MSNM-S. En este trabajo, se extraen manualmente en todos los experimentos[9]

[9]Para más información sobre la configuración de los sensores MSNM, consultar el artículo original [18].

Todos los experimentos incluyen una parte de tráfico normal y un periodo de ataques. En este trabajo (por tratarse de un prototipo), las pruebas se limitan a ataques de DoS , pues es uno de los ataques más frecuentes en entornos IoT. En cada prueba se lanza al menos una ráfaga de ataques de varios minutos de duración con el objetivo de comprobar si son detectados por el prototipo. Los detalles de estos ataques se resumen en la Tabla III.

IV-A. Experimento 1 (E01)

Antes de comenzar con el primer experimento, se lleva a cabo una captura de datos en condiciones normales de operación para comprobar el funcionamiento del escenario propuesto, así como de los sensores de MSNM configurados. A continuación, se configura el experimento según lo descrito en la Tabla II (sensores MSNM) y Tabla III (ataques).

Para analizar los resultados, se ha representado la evolución de los estadísticos monitorizados en cada uno de los puntos desplegados en la Figura 2 (los triángulos invertidos azules se corresponden con el Q-st, mientras que los círculos naranjas representan el D-st). También se han representado los límites de control para poder identificar cuando existen anomalías. En concreto, para el Q-st (UCLq) se representan con una línea verde discontinua, mientras que para el D-st (UCLd) se representan con una línea roja continua. Como la calibración es dinámica, no ha sido necesario capturar datos NOC antes de la monitorización, sino que se adaptan los límites conforme pasa el tiempo y se reciben nuevas observaciones.

En la Figura 2 (a) y (b) se representan los estadísticos y sus correspondientes límites de control para los sensores de los dispositivos D1 y D2 (izquierda y derecha, respectivamente). Se puede observar que los límites de control se auto ajustaron después de media hora de ejecución (el modelo se recalcula cada media hora, B=30), quedando prácticamente coincidentes tanto para el modelo (UCLd) como para el residuo (UCLq). Inicialmente (durante los primeros treinta minutos) el Q-st se encontraba por encima de su límite de control, lo cual es normal, ya que aún no se había ajustado. Tras el ajuste, se detectan varias anomalías tanto en el modelo como en el residuo[10]. Entre las 13:10 y las 13:20 se observa un incremento desmesurado del valor de ambos estadísticos que es asignable al ataque de DoS introducido durante la experimentación. Una vez finaliza el ataque, ambos estadísticos quedan alterados durante unos minutos, pero luego se estabilizan. Además, se puede observar que el ataque afecta a D1, pues los estadísticos en D2 están, en general, bajo control, mientras que en D1 se observa una elevación claramente anómala durante el periodo del ataque de DoS.

Finalmente, en la Figura 2 (c) y (d) se representan las gráficas con los estadísticos para el *broker* y la base de datos, respectivamente. De nuevo, entre las 13:10 y las 13:20 aumenta el valor de los estadísticos, coincidiendo con el ataque de DoS introducido durante la experimentación. Se puede observar que la magnitud de los valores es superior respecto a los observados en los dispositivos. Cuanto más superior es el nivel en la jerarquía, más se magnifican los estadísticos. En este caso, el *broker* se encuentra en el nivel intermedio (nivel 2) y la base de datos en el último (nivel 3). La amplificación de los valores de los estadísticos de monitorización en los niveles superiores de la jerarquía es un comportamiento inherente de la metodología propuesta. Este comportamiento puede ser útil para la detección de ataques no tan fácilmente discernibles y que se mimetizan con el tráfico normal. Algunos ejemplos de ataques son aquellos que implican comportamientos derivados del control de redes de *bots* o exfiltración de datos.

[10]Sería necesario diagnosticar la causa de estas anomalías para comprender si se debe a un mal ajuste del modelo o si realmente estaban causadas por factores externos. No se ha llevado a cabo diagnóstico porque queda fuera del ámbito de este trabajo (probar el funcionamiento de los sensores de MSNM en un entorno realista de IoT). En trabajos futuros se realizará la diagnosis.

IV-B. Experimento 2 (E02)

En este experimento se considera que la calibración inicial del modelo PCA será la misma durante todo el experimento, es decir, estática. Para ello, primero se captura tráfico del escenario IoT en condicionales normales, es decir, sin la presencia de ataques. Dicho tráfico se procesa mediante la aproximación FaaC, convirtiéndose así en la matriz necesaria para calibrar. Este proceso se repite para cada uno de los sensores MSNM desplegados en el sistema. En este experimento se despliegan únicamente tres sensores: uno por cada dispositivo IoT (D1 y D2) y otro en el *broker*. Tanto D1 como D2, envían sus estadísticos al *broker*, donde son agregados junto al tráfico de red que éste observa. Así mismo, se llevan a cabo dos ráfagas de ataque DoS, iniciadas desde D1, que se presupone comprometido, hacia la base de datos.

En la Figura 3 se muestran los diagramas de monitorización obtenidos por los diferentes sensores de MSNM. Tanto en D1 como en el *broker* se pueden observar claramente las dos ráfagas de ataque, no apreciándose cambios significativos en los estadísticos de D2. Esto indica que el ataque se generó en D1. Se puede comprobar que, de nuevo, existe una diferencia de escala entre los estadísticos obtenidos para D1 y D2 y los del *broker*. Si se comparan la Figura 2 y la Figura 3, se puede observar cómo los límites de control se adaptan al tráfico de red en E01, mientras que en E02 permanecen constantes. Llama la atención que los estadísticos son más inestables durante E01. En general, ambas configuraciones funcionan correctamente, aunque, para este escenario, la configuración con modelo estático (E02) presenta unos resultados más ”limpios”. Esto podría deberse a que el tráfico de esta red IoT no es ciclo-estacionario. Como trabajo futuro, queda pendiente identificar las causas de esta inestabilidad, así como identificar cuál es la mejor configuración de MSNM-S para entornos IoT.

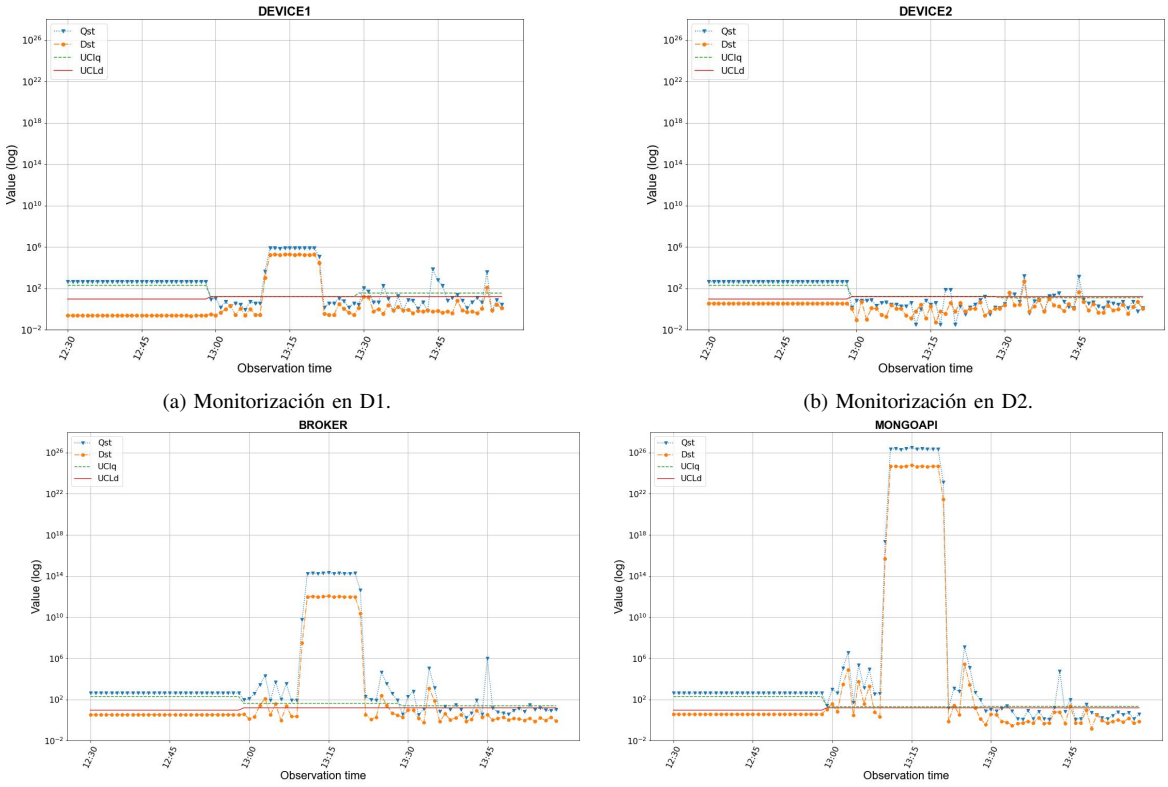
V. CONCLUSIONES

Este prototipo es fruto de la colaboración activa entre la Universidad de Granada (UGR) y Fidesol. Esta colaboración se ha materializado en forma de transferencia de conocimiento del sensor de MSNM (desarrollado por la UGR), cuya integración y puesta en marcha en Atenea Lab se ha llevado a cabo por el equipo de Investigación y Desarrollo de Fidesol.

La experimentación llevada a cabo en ambos centros ha hecho posible evaluar el rendimiento del prototipo en un entorno IoT realista de forma satisfactoria, siendo la primera vez que se que aplica y valida MSNM-S en ecosistemas IoT. Tras llevar a cabo experimentos en distintas condiciones, se ha comprobado que: *i)* Atenea Lab funciona correctamente, *ii)* el prototipo es capaz de detectar ataques contra la disponibilidad

Tabla III: Ataques DoS llevados a cabo para cada experimento

Id. exp.	Inicio captura	Fin captura	Inicio ataque	Fin ataque	D. Afectado
E01	07/12/2022 12 : 30	07/12/2022 14 : 00	07/12/2022 13 : 10	07/12/2022 13 : 20	D1
E02	17/04/2023 18 : 30	17/04/2023 21 : 20	17/04/2023 20 : 36 17/04/2023 21 : 11	17/04/2023 20 : 41 17/04/2023 21 : 13	D1



(a) Monitorización en D1.

(b) Monitorización en D2.

(c) Monitorización en el *broker*.

(d) Monitorización en el Base de Datos.

Figura 2: En la parte superior de la figura se representa la evolución de los estadísticos para (a) el dispositivo 1 y (b) el dispositivo 2, mientras que en la parte inferior de la figura se muestran los estadísticos para (c) el *broker* y (d) la base de datos.

de sistemas IoT en tiempo real y *iii)* existen artefactos que necesitan una investigación más profunda para poder avanzar en la evolución del prototipo hacia un TRL superior.

Ambos centros tienen previsto seguir colaborando para el estudio y corrección de estos artefactos, la validación del prototipo con distintos tipos de ataque[11], la inclusión de nuevos modos de fusión de datos en los distintos niveles de la jerarquía, la validación de la escalabilidad del prototipo y la inclusión de una interfaz gráfica de usuario que permita una interacción más sencilla, dinámica y comprensible por parte de los analistas de seguridad. De hecho, MSNM-S forma parte de la tecnología base de la que parte el proyecto SOCIABLE (SOC basado en IA interpretable), presentado a la CPI de INCIBE[12], para su desarrollo de forma colaborativa entre Fidesol y los grupos de investigación NESG y CoDas.

[11]Nótese que el enfoque de detección de anomalías, por su naturaleza no supervisada, permite detectar varios tipos de ataque, incluso *zero-day*)

[12]Convocatoria 3-Reto 7. https://www.incibe.es/industria-cpi/cpi-tercera-convocatoria

AGRADECIMIENTOS

Este trabajo está financiado en parte por las Ayudas Cervera para Centros Tecnológicos del Centro Español para el Desarrollo de Tecnología Industrial (CDTI) en el marco del proyecto EGIDA (CER-20191012) y por el Ministerio de Ciencia e Innovación (MICIN) MICIN/AEI/10.13039/501100011033, bajo los proyectos PID2020-113462RB-I00 y PID2020-114495RB-I00, así como los proyectos PPJIA2022-51 y PPJIA2022-52 de ayudas del plan propio de la UGR.

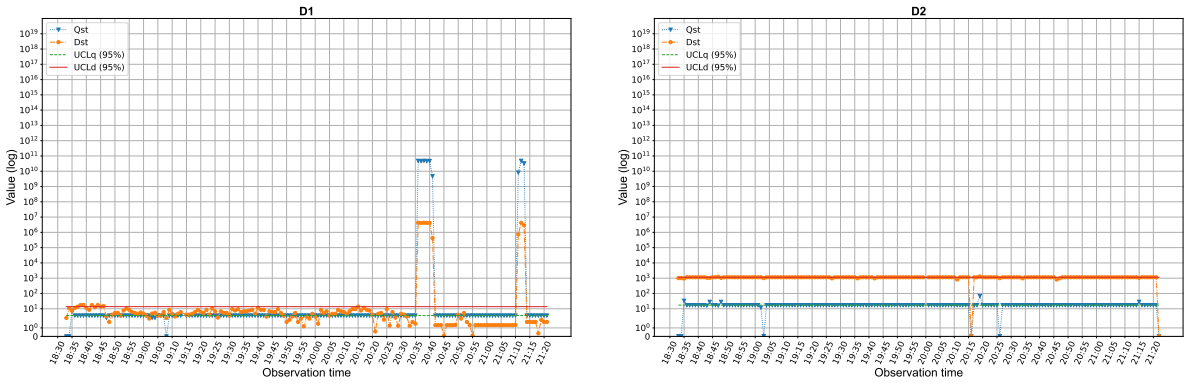
REFERENCIAS

[1] ASAJA. El sector agrario constituirá en 2021 la principal palanca para la reactivación económica de Andalucía. [accedido el 03/03/2023]. [Online]. Available: https://bit.ly/3JbWhwi

[2] Pwc, "El futuro del sector agrícola español," Pwc, Tech. Rep., [Accedido el 03/03/2023]. [Online]. Available: https://www.pwc.es/es/publicaciones/assets/informe-sector-agricola-espanol.pdf

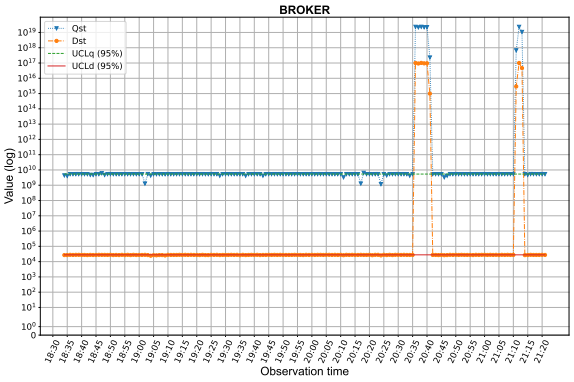
[3] P. P. Ray, "Internet of Things for Smart Agriculture: Technologies, Practices and Future Direction," *IOS Press*, vol. 9, pp. 395–420, 2017.

[4] A. Tzounis, N. Katsoulas, T. Bartzanas, and C. Kittas, "Internet of Things in agriculture, recent advances and future challenges," *Biosystems Engineering*, vol. 164, pp. 31–48, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1537511017302544



(a) Monitorización en el D1.

(b) Monitorización en el D2.



(c) Monitorización en el *broker*.

Figura 3: En la parte superior de la figura se representa la evolución de los estadísticos para (a) el dispositivo 1 y (b) el dispositivo 2, mientras que en la parte inferior de la figura se muestran los estadísticos para (c) el *broker*.

[5] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, Opportunities, and Directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724–4734, 2018.

[6] M. A. Ferrag, L. Shu, X. Yang, A. Derhab, and L. Maglaras, "Security and Privacy for Green IoT-Based Agriculture: Review, Blockchain Solutions, and Challenges," *IEEE Access*, vol. 8, pp. 32 031–32 053, 2020.

[7] Gartner, "Gartner Identifies the Top Seven Security and Risk Management Trends for 2019," Gartner, Tech. Rep., [accedido el 03/03/2023]. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2019-03-05-gartner-identifies-the-top-seven-security-and-risk-ma

[8] I. Porro-Sáez, "Iot: protocolos de comunicación, ataques y recomendaciones," INCIBE (Ministerio de Economía y Empresa), https://www.incibe-cert.es/blog/iot-protocolos-comunicacion-ataques-y-recomendaciones, Tech. Rep., 2019, [Online, accessed on 15/04/2020].

[9] INCIBE, "Seguridad en redes wifi: una guía de aproximación para el empresario," INCIBE (Ministerio de Economía y Empresa), https://cutt.ly/7ySrhmy, Tech. Rep., 2019, [Online, accessed on 14/04/2020].

[10] D. Dragomir, L. Gheorghe, S. Costea, and A. Radovici, "A Survey on Secure Communication Protocols for IoT Systems," in *2016 International Workshop on Secure Internet of Things (SIoT)*, 2016, pp. 47–62.

[11] P. G. S. P. I. R. Grammatikis and I. D. Moscholios, "Securing the Internet of Things: Challenges, threats and solutions," vol. 5, pp. 41–70.

[12] K. T. Nguyen, M. Laurent, and N. Oualha, "Survey on secure communication protocols for the internet of things," *Ad Hoc Networks*, vol. 32, pp. 17 – 31, 2015, internet of Things security and privacy: design methods and optimization. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570870515000141

[13] Eleven-Paths, "Informe sobre el estado de la seguridad 2020 H2," Eleven Paths. Telefonica, Tech. Rep., [Online, accessed on 08/02/2021]. [Online]. Available: tinyurl.com/2jsodhs3

[14] P. Nair, "Hackers Breached Israeli Water Reservoir HMI System," Bank Info Security, Tech. Rep., [Accedido el 03/03/2023]]. [Online]. Available: tinyurl.com/4qwt58c4

[15] F. Robles and N. Perlroth, "'Dangerous Stuff': Hackers Tried to Poison Water Supply of Florida Town," [Online, accessed on 11/02/2021]. [Online]. Available: tinyurl.com/2hdh6tr9

[16] J. Cox, "Agriculture Industry on Alert After String of Cyber Attacks," The Bakersfield Californian, Tech. Rep., [Online, accessed on 31/08/2022]. [Online]. Available: https://bit.ly/3RmKvjJ

[17] J. Camacho, A. Pérez-Villegas, P. García-Teodoro, and G. Maciá-Fernández, "PCA-based Multivariate Statistical Network Monitoring for Anomaly Detection," *Computers & Security*, vol. 59, pp. 118–137, June 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167404816300116

[18] R. Magán-Carrión, J. Camacho, G. Maciá-Fernández, and Ruíz-Zafra, "Multivariate Statistical Network Monitoring–Sensor: An effective tool for real-time monitoring and anomaly detection in complex networks and systems," *International Journal of Distributed Sensor Networks*, vol. 16, no. 5, p. 1550147720921309, May 2020.

[19] T. Kourti and J. F. MacGregor, "Multivariate SPC methods for process and product monitoring," vol. 28, no. 4, pp. 409–428.

[20] M. Fuentes-García, "Multivariate Statistical Network Monitoring for Network Security based on Principal Component Analysis."

[21] J. Camacho, "Visualizing big data with compressed score plots: Approach and research challenges," *Chemometrics and Intelligent Laboratory Systems*, vol. 135, pp. 110–125, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016974391400080X

[22] G.-T. P. Camacho, J. and G. Maciá-Fernández, "Traffic Monitoring and Diagnosis with Multivariate Statistical Network Monitoring: A Case Study," in *IEEE Security Privacy International Workshop on Traffic Measurements for Cybersecurity (WTMC2017)*, pp. 241–246.

[23] J. Camacho, J. M. García-Giménez, N. M. Fuentes-García, and G. Maciá-Fernández, "Multivariate Big Data Analysis for intrusion detection: 5 steps from the haystack to the needle," vol. 87, p. 101603.

[24] C. J. G.-T. P. Maciá-Fernández, G. and Rodríguez-Gómez, "Hierarchical PCA-Based Multivariate Statistical Network Monitoring for Anomaly Detection."