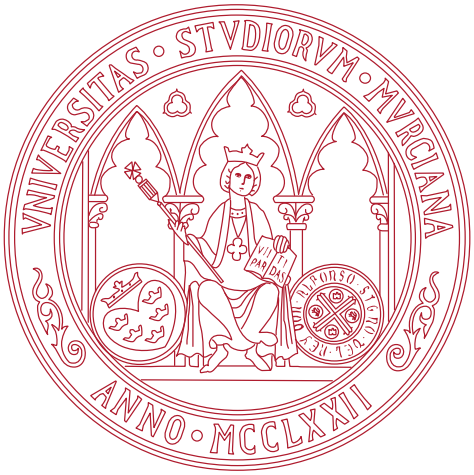
UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

Detección de Bonets y Ransomware en Redes de Datos mediante Técnicas de Aprendizaje Automático

D. Lorenzo Fernández Maimó

2019



UNIVERSIDAD DE MURCIA

FACULTAD DE INFORMÁTICA

Detección de Botnets y Ransomware en Redes de Datos Mediante Técnicas de Aprendizaje Automático

Autor

Lorenzo Fernández Maimó

Director de Tesis

Dr. Félix Jesús García Clemente

Murcia, 2019

La presente Tesis Doctoral es un compendio de los siguientes artículos publicados, siendo el doctorando el autor principal en todos ellos:

1. Lorenzo Fernández Maimó, Ángel Luis Perales Gómez, Félix J. García Clemente, Manuel Gil Pérez, Gregorio Martínez Pérez. "A Self-Adaptive Deep Learning-Based System for Anomaly Detection in 5G Networks", IEEE Access, vol. 6, pp. 7700–7712, febrero 2018. https://doi.org/10.1109/ACCESS.2018.2803446

2. Lorenzo Fernández Maimó, Alberto Huertas Celdrán, Manuel Gil Pérez, Félix J. García Clemente, Gregorio Martínez Pérez. "Dynamic management of a deep learning-based anomaly detection system for 5G networks", Journal of Ambient Intelligence and Humanized Computing, Pub. online: mayo 2018. https://doi.org/10.1007/s12652-018-0813-4

3. Lorenzo Fernández Maimó, Alberto Huertas Celdrán, Ángel Luis Peráles Gómez, Félix J. García Clemente, James Weimer, Insup Lee. "Intelligent and Dynamic Ransomware Spread Detection and Mitigation in Integrated Clinical Environments", Sensors, vol. 19(5) 1114, marzo 2019. https://doi.org/10.3390/s19051114

Índice

Agradecimientos iii

Resumen 1

I Introducción 1
II Metodología 6
III Resultados 8
IV Conclusiones y Trabajo Futuro 11

Publicaciones que componen la tesis doctoral 14

1 A self-adaptive deep learning-based system for anomaly detection in 5G networks 19

2 Dynamic management of a deep learning-based anomaly detection system for 5G networks 21

3 Intelligent and dynamic ransomware spread detection and mitigation in ICE 23

Referencias 27

Agradecimientos

Agradecimientos

En primer lugar me gustaría dar las gracias a mi familia al completo por estar ahí en los buenos y malos momentos, y especialmente a mi madre y abuelos por los valores que me han transmitido y que han sido en gran parte los causantes de estar hoy aquí. Gracias a José Ignacio, amoroso y paciente durante estos años, que me ha estado animando en los estresantes últimos días de cada plazo de entrega de los artículos, y acompañando en tantos fines de semana trabajando en casa; te quiero. Gracias a mis amigos por ser uno de los grandes pilares que sustentan mi vida. Gracias a todos.

Quisiera también agradecer a mis compañeros de departamento el apoyo moral recibido, y el buen clima de trabajo del que hemos disfrutado en estos años, especialmente a los amigos de MOVI por los buenos momentos, tanto en las reuniones como en la comida semanal. Mención especial merecen Pedro y Alberto R., que han sido y siguen siendo para mí un referente por sus vastos conocimientos a los que tantas veces he recurrido, por sus siempre buenos consejos y su apoyo permanente. Gracias por vuestra amistad.

Gracias a Alberto H. y Angel Luis, compañeros de fatigas en este periplo, y con los que tantos ratos de trabajo a horas intempestivas he compartido. Gracias por estar siempre ahí para echar una mano. Mis agradecimientos también a Gregorio y Manuel por hacerme sentir como uno más de su grupo de investigación.

Finalmente, quisiera agradecer a mi director, Félix, el haberme despertado de mi letargo investigador, el tener esa santa paciencia para contrarrestar mi visión pesimista en ciertos momentos, y el haber sabido focalizar mi habitualmente disperso esfuerzo en una dirección concreta, dando como resultado esta tesis que hoy tienes en tus manos.

Resumen

I Introducción

En el campo de la ciberseguridad, profesionales e investigadores han diseñado a lo largo de los años una variedad de sistemas de ciberdefensa con el fin de proteger las organizaciones de atacantes maliciosos. Estos sistemas se enfrentan a amenazas tales como virus, troyanos, gusanos y botnets entre otras. Las soluciones existentes basadas en Sistemas de Detección de Intrusiones (IDS) incluyen enfoques proactivos para anticipar vulnerabilidades en sistemas informáticos y así poder ejecutar acciones de mitigación. Sin embargo, con los años el número de amenazas ha aumentado enormemente, sobre todo por la aparición de entornos de desarrollo de malware capaces de generar casi automáticamente diferentes versiones de un mismo virus, haciendo que cualquier aficionado pueda producir su propia variación. Esta proliferación de malware hace que las bases de datos de reglas usadas por los IDS sean cada vez mayores, incrementando, por tanto, el tiempo de cómputo necesario para dicha detección.

En particular, estos sistemas de detección presentan en la actualidad tres puntos débiles que merece la pena destacar: a) el equilibrio necesario entre la efectividad de la detección de amenazas y la velocidad a la que se pueden examinar los datos recogidos en las redes de datos modernas; b) la necesidad de poder detectar dichas amenazas incluso cuando los datos viajan cifrados; y c) la dificultad de detectar nuevas versiones de malware aunque sean variaciones de una familia ya conocida.

El primero viene motivado por las velocidades de transferencia que las redes de comunicaciones están alcanzando (por encima de los 10 Gbps), con volúmenes crecientes de intercambio de datos. A esto hay que añadir la nueva tecnología móvil de quinta generación (5G), que promete proporcionar una latencia y una velocidad de transferencia nunca vistas a las redes inalámbricas, permitiendo una expansión sin precedentes del Internet de las Cosas —*Internet of Things* (IoT)—. Todo esto hará realmente difícil capturar y analizar cada paquete que circule por la red, provocando que los procedimientos de detección actuales se queden obsoletos si no somos capaces de adaptarlos adecuadamente. Precisamente por la ubicuidad del acceso a Internet, en parte gracias a las tecnologías inalámbricas, y el enorme crecimiento de la IoT, existen millones de dispositivos con vulnerabilidades que pueden ser usados para formar botnets.

Una botnet es un conjunto de dispositivos conectados a Internet que un atacante infectó previamente con un software que le permite manejarlos de forma remota para realizar todo tipo de acciones como, por ejemplo, ataques distribuidos de denegación de servicio, robo de información sensible o crítica como cuentas bancarias y datos personales, o incluso robo de

ciclos de CPU para minar criptomonedas. Las botnets se han convertido en uno de los grandes problemas de seguridad actuales y futuros en las redes de datos; baste mencionar que, según el reciente Nokia Threat Inteligence Report, en 2018 las botnets basadas en IoT supusieron el 78% de los ataques por malware detectados y el 16% de los dispositivos IoT infectados [1].

Por tanto en el contexto de las comunicaciones 5G y el IoT, la detección de malware se convierte en un reto debido a la cada vez mayor diversidad de dicho malware, cuyas reglas hay que aplicar a cada paquete, y a que las tasas de transferencia tan altas junto con el gran volumen de datos que se mueve, dejan poco tiempo para examinar cada paquete que circula por la red. Cuando evaluamos el volumen de paquetes que las actuales herramientas de inspección profunda de paquetes pueden gestionar, nos encontramos con que la conocida Snort soporta redes cableadas de hasta 1 Gbps, empezando a descartar paquetes debido a sobrecarga a partir de 1,5 Gbps [2]. Esto ha provocado la aparición de soluciones hardware basadas en matrices de puertas programables (FPGA) [3] o circuitos integrados de propósito específico (ASIC), que permiten trabajar con velocidades de hasta 7,2 Gbps [4]. Aún así, estas velocidades quedan lejos de las que nos esperan en el futuro cercano. Debido en parte a esto, las soluciones de detección basadas en IDS han tenido que evolucionar y pasar de analizar paquetes de red a analizar flujos de tráfico de red por medio de novedosas técnicas basadas en la inteligencia artificial [5]. Por ejemplo, un modelo de red neuronal basada en bloques usada en un IDS basado en anomalías en flujos pudo trabajar con tráfico a 22 Gbps usando FPGAs [6]. Una revisión completa de soluciones para clasificar rápidamente flujos de red y detectar ataques o código malicioso puede encontrarse en [7].

Con respecto al segundo punto débil, un número cada vez mayor de malware cifra sus comunicaciones [8], lo que imposibilita el examen profundo de los paquetes y hace ineficaces las herramientas habituales de detección. A esto hay que añadir el progresivo incremento de la cantidad de tráfico cifrado en las comunicaciones del día a día, y la obligatoriedad de usar dicho cifrado en entornos donde la privacidad es crítica, como por ejemplo entornos médicos. Estos entornos están incorporando en sus proyectos para la evolución del hospital del futuro, un creciente número dispositivos médicos interoperables con el fin de poder llegar a implementar procesos de ciclo cerrado (monitorización, análisis, toma de decisión y reacción o aplicación de un tratamiento), mientras que al mismo tiempo han estado sufriendo en los últimos años un número cada vez mayor de ataques por malware exitosos, demostrando que los sistemas de detección actuales son poco efectivos. Un malware especialmente problemático ha sido el denominado ransomware, que consiste en infectar a uno de los dispositivos de una red gracias a una vulnerabilidad o fallo humano, y llegar al resto de dispositivos por medio de una propagación horizontal basada habitualmente en vulnerabilidades del sistema, tras lo cual cifra todos los datos que contienen sus discos duros y carpetas compartidas y exige una cantidad de dinero para proporcionar la clave de descifrado. Para apreciar el peligro potencial que supone el ransomware, baste mencionar los ataques sufridos por los hospitales del servicio de salud de Reino Unido en 2017, que llegaron a tener que cerrar servicios enteros, enviar pacientes a otros hospitales e incluso posponer intervenciones quirúrgicas [9].

En este contexto de tráfico cifrado, las soluciones basadas en inspección profunda tampoco son aplicables, y la mayoría de soluciones enfocadas a IDS que manejan tráfico cifrado se basan en identificar ciertos patrones básicos como pueden ser escaneo de puertos, o ataques por fuerza bruta [10]. Hay propuestas basadas en aprendizaje automático que usan datos calculados a partir de un flujo [11], e incluso ofrecen propuestas imaginativas que usan redes convolucionales para tratar el flujo como si fuera una imagen [12]. Sin embargo, si no podemos acceder a la carga útil por ir cifrada, un único flujo no proporciona información suficiente como para conseguir una detección precisa.

Finalmente, el tercer punto débil de nuestro interés viene motivado por la proliferación de

nuevo malware, generalmente derivado de versiones existentes a las que se les cambian características como la arquitectura sobre la que se ejecutan o el método de cifrado [13]. Los IDS tienen dificultades en identificar estas variaciones al trabajar normalmente examinando el tráfico por medio de reglas, lo que imposibilita su detección temprana. En el caso de las botnets y ransomware, ambos permiten fácilmente la generación de nuevas versiones. Como muestra de ello, un informe de Heimdal Security de 2016 [14] concluye que nueve de las diez botnets de filtrado de datos más peligrosas son variaciones de la botnet Zeus. Igualmente en ransomware encontramos gran diversidad de miembros de la misma familia, por ejemplo, del ransomware Petya se derivan NotPetya, ExPetr o PetrWrap entre otros.

Botnets y ransomware tienen en común que generan tráfico de red siguiendo unos patrones característicos. En el caso de las botnets, lo habitual es que tengan un mecanismo de mando y control —*Command and Control (C&C)*— mediante el cual cada dispositivo infectado se comunica con el ordenador del atacante periódicamente para recibir órdenes. En el caso del ransomware, sus patrones de tráfico provienen de su afán por propagarse horizontalmente para maximizar el daño y con ello aumentar la probabilidad de pago del rescate, de la comunicación con un servidor central para la obtención de las claves de cifrado, y del tráfico necesario para el cifrado de carpetas compartidas en red de los equipos infectados. Estos patrones pueden interpretarse como anomalías en el tráfico normal de la red.

Detección de anomalías en redes de datos

Una anomalía puede definirse como un patrón que no se ajusta al comportamiento esperado o normal, lo que implica que aparece muy poco frecuentemente. Precisamente por basarse en el concepto de normalidad, que de por sí no es fácil de definir, el problema dista mucho de ser sencillo, existiendo principalmente tres grandes categorías de anomalías [15]:

- Puntuales. Es la forma más simple de anomalía y es donde se centra la mayoría de la investigación en este campo. Una muestra de un grupo de datos que se considera anómala respecto al resto, es un ejemplo de este tipo.
- Contextuales. Una instancia es anómala en un contexto determinado, pero no en otro. Por ejemplo, una temperatura de 5 grados es anómala o no dependiendo de la estación.
- Colectivas. Una colección de instancias puede considerarse una anomalía con respecto a un conjunto de datos si, aunque cada una de las instancias no supone una anomalía, la aparición de todas ellas como una colección sí lo es. Un ligero alargamiento de parte de la onda de un electrocardiograma puede ser un ejemplo de este tipo de anomalía. Los valores pertenecen al rango normal, pero la secuencia de valores en sí misma constituye la anomalía.

Las técnicas de detección de anomalías basadas en aprendizaje automático actúan como clasificadores capaces de distinguir instancias anómalas y normales, pudiéndose encontrar aproximaciones basadas en aprendizaje supervisado, semi-supervisado y no supervisado. Dentro de estas categorías, es de esperar que, si existe un conjunto de datos etiquetado adecuadamente, el enfoque supervisado dé los mejores resultados al tener información más completa. Sin embargo este enfoque tiene como principales retos la dificultad de conseguir un conjunto de datos representativo y el hecho de que las anomalías suelen ser órdenes de magnitud menos numerosas que los casos normales, con el consiguiente desequilibrio del conjunto de datos que complica la tarea de clasificación.

Las anomalías en el tráfico de red pueden pertenecer a cualquiera de los tres tipos anteriores. Por poner un ejemplo de cada uno, una anomalía puntual podría ser dirigir un paquete a un

puerto sospechoso; en el tráfico entre dos dispositivos que intercambian paquetes pequeños, una anomalía contextual sería la aparición de un paquete de gran tamaño, ya que no sería considerado una anomalía si los dispositivos implicados fuesen otros de la misma red que sí usaran ese tamaño de paquete; por último, en un contexto en el que se emiten paquetes en ráfagas de una cierta duración conocida, un cambio en esa duración puede considerarse una anomalía colectiva. En esta investigación se utiliza la agregación de flujos a lo largo del tiempo para poder aplicar métodos de detección de anomalías puntuales a anomalías contextuales y colectivas.

Uso de flujos de red en la detección de anomalías

Utilizar los flujos para detectar dichas anomalías tiene múltiples ventajas, entre las que se encuentran las siguientes: no se necesita acceder a la carga útil del paquete, por lo que es aplicable a tráfico cifrado; reduce en órdenes de magnitud el volumen de datos a analizar, por lo que se puede aplicar a entornos con redes de alta velocidad; por último, respetan la privacidad del usuario. El tema de la privacidad es algo crucial a la hora de obtener permiso de las organizaciones para capturar la ingente cantidad de tráfico que algunos de los algoritmos de aprendizaje automático más recientes necesitan para su entrenamiento. Al trabajar con flujos es más sencillo garantizar el anonimato y la privacidad de los usuarios, puesto que los administradores saben que sólo se precisarán las cabeceras de los paquetes. Estas razones los hacen atractivos para esta investigación.

Por supuesto, trabajar con flujos también tiene inconvenientes. El principal es la pérdida de información que implica al perder el contenido de los paquetes y limitarse a obtener una información agregada de la secuencia de paquetes que componen el flujo. Sin embargo, esta investigación pretende mostrar que aunque un único flujo puede no ser suficiente para extraer información compleja del patrón del tráfico, una secuencia de flujos, tratada adecuadamente, puede contener suficiente información como para que un algoritmo de aprendizaje automático pueda diferenciar los patrones anómalos de los normales. De esta manera cumpliríamos las restricciones de los escenarios propuestos, al realizar la detección sin necesitar examinar un alto volumen de datos ni preocuparnos por el hecho de que los datos viajen cifrados.

Infraestructura para la detección de anomalías

Con respecto a la forma de integrar las propuestas de esta investigación en la infraestructura de comunicaciones, se precisa de unos mecanismos que faciliten tareas como la actualización de modelos, adaptación a las circunstancias del entorno, e incluso mitigación automática e inteligente tras la detección de las amenazas. Para ello se decidió utilizar virtualización de funciones de red (NFV) que proporciona flexibilidad y dinamismo en la infraestructura al separar la capa hardware de la software, y redes definidas por software (SDN) que nos aporta control de la comunicaciones en tiempo real y bajo demanda. La unión de ambas permite, entre otras posibilidades, integrar de una forma natural el proceso de adquisición de datos a partir de los flujos de red, y la optimización de recursos durante la mitigación de ataques de forma automática, en tiempo real y bajo demanda. Por poner un ejemplo, en el caso del escenario 5G, se podrían reasignar recursos para adaptarse a un mayor volumen de tráfico que requiera de una GPU para conseguir más rendimiento; mientras que en escenario clínico integrado, permiten sustituir en tiempo real un dispositivo virtual comprometido.

En la literatura existen propuestas que emplean algunas de las posibilidades que ofrecen NFV/SDN para la detección y mitigación de ransomware [16, 17, 11]. Sin embargo esta investigación es novedosa por estar basada en la combinación NFV/SDN como medio para recolectar los datos de los flujos de forma transparente, permitir adaptarse a las circunstancias del tráfico, y al mismo tiempo detectar y mitigar amenazas de manera inteligente y dinámica. Este

planteamiento se ha aplicado a dos entornos diferentes: uno de comunicaciones inalámbricas basado en 5G siguiendo el paradigma de *Edge Computing*, y otro consistente en un entorno clínico integrado (ICE). El primer entorno tiene como principal característica una enorme densidad de tráfico, que exige un tiempo de análisis del tráfico muy bajo para evitar descartar información y está centrado en anomalías provocadas por botnets, que tienen la peculiaridad de permanecer sin causar daño durante largos periodos de tiempo, lo que ayuda a su detección y posterior mitigación. El segundo tiene como restricción principal el tiempo de detección y mitigación de la anomalía provocada por el ransomware, que se activa en un corto espacio de tiempo y por tanto, la detección y mitigación deben producirse antes de que comience el cifrado de los ficheros.

Objetivo

El principal objetivo de esta tesis consiste en investigar la forma de aplicar métodos de aprendizaje automático a la detección de anomalías en redes de datos con restricciones. En un primer caso, la restricción la impone la imposibilidad de analizar la carga útil de todos los paquetes por el volumen de tráfico circulante; y en un segundo caso, la restricción consiste en tener que trabajar con tráfico cifrado y el corto tiempo para la detección y mitigación. La hipótesis que esta tesis plantea tras analizar las propuestas para clasificar el tráfico de red basadas en flujos existentes en la literatura, es que un flujo por sí solo, sin acceso a la carga útil de los paquetes, no aporta suficiente información; y se propone estudiar si un contexto para ese flujo, formado por los flujos recibidos previamente durante un periodo de tiempo, permitiría una detección más precisa de anomalías en patrones de tráfico complejos, para lo cual será necesario emplear métodos de aprendizaje automático de detección de anomalías, tanto clásicos como profundos. A la vez defiende que la evaluación del tráfico podrá hacerse a la velocidad que las exigentes redes 5G obligan, y que el tiempo de detección permitirá mitigar un ataque por ransomware antes de que se propague. Todo esto de forma dinámica, inteligente, en tiempo real, e integrado dentro de una arquitectura adecuada para cada entorno. Para alcanzar este objetivo, se han llevado a cabo una serie de acciones específicas detalladas a continuación:

1. Estudiar el concepto de anomalía y las principales técnicas de aprendizaje automático propuestas en la literatura para su detección, junto con sus requerimientos computacionales y de almacenamiento. En particular esto incluye el estudio del estado del arte de los sistemas de detección de anomalías en tráfico de red, así como su necesidad de inspeccionar la carga útil de los paquetes.

2. Determinar al menos dos entornos representativos donde poder comprobar el rendimiento de las propuestas, y que nos permitan obtener tiempos de ejecución y despliegue.

3. Diseñar un mecanismo de generación de vectores de características a partir de flujos extraídos del tráfico capturado, que incluya información acerca del comportamiento de dicho tráfico durante un periodo de tiempo, a la vez que permite determinar o acotar el causante del comportamiento anómalo sin vulnerar la privacidad del usuario.

4. Integrar este mecanismo en una arquitectura basada en NFV/SDN, específica para cada entorno, con un ciclo de vida que permita tanto una adaptación dinámica de los modelos a las nuevas circunstancias de la red, como facilitar la mitigación de las amenazas.

5. Analizar los conjuntos de datos públicos existentes con tráfico de red real conteniendo tráfico normal y tráfico generado tanto por botnets como por ransomware para su posible uso en este trabajo. Evaluar asimismo la posibilidad de generar un nuevo conjunto de datos.

6. Evaluar las propuestas en cada uno de los entornos seleccionados por medio de los conjuntos de datos obtenidos, en términos de capacidad de detección, así como tiempos de detección y mitigación.

II Metodología

En esta tesis doctoral se ha seguido una metodología científica basada en el estudio del estado del arte de los sistemas de detección de anomalías basados en aprendizaje automático en el contexto de las redes de datos. Siguiendo el primer objetivo definido en la sección anterior, se comenzó con el estudio del concepto de anomalía, especialmente en el contexto de las redes de datos, y de las principales técnicas de aprendizaje automático utilizadas para su detección, haciendo énfasis en aquellas que no precisan de la carga útil del paquete. Puesto que es razonable suponer que se puede capturar el tráfico considerado normal, ya se tendría al menos la clase normal etiquetada, y por tanto se centró el estudio en los métodos semi-supervisados y supervisados. En un primer momento se podría pensar que un enfoque supervisado no tiene sentido, puesto que no se podrán detectar anomalías desconocidas; sin embargo, nuestra hipótesis es que con la suficiente variedad de muestras, hay comportamientos compartidos entre familias de malware del mismo tipo, tanto en botnets como en ransomware, que pueden ser aprendidos, resultando en la detección de ejemplos desconocidos por su asociación con uno de los conocidos mediante un clasificador probabilístico. Además, gracias a recientes propuestas [11], la obtención y etiquetado del conjunto de datos se ven como una opción viable.

Durante la investigación realizada, y siguiendo el segundo objetivo enumerado, se identificaron dos escenarios donde la detección de anomalías suponía un reto por las restricciones que cada uno imponía. El primero fue el entorno de las comunicaciones móviles 5G, donde las altas tasas de transferencia y el gran volumen de datos hace extremadamente difícil analizar cada paquete del tráfico circulante, lo que puede retrasar la detección. Además, dadas las fluctuaciones del tráfico en 5G, esta arquitectura de comunicación necesita adaptar dinámicamente los recursos dedicados a la detección de anomalías para no malgastarlos. Tras analizar las anomalías más frecuentes en estas redes, se consideró que las provocadas por botnets eran unas de las más representativas. El segundo escenario se decidió que fuera el de los entornos clínicos integrados (ICE), donde la mayoría del tráfico de red circula cifrado, con el añadido de que en un entorno tan crítico se hace imprescindible una detección y mitigación lo más rápida posible, máxime cuando en los últimos años varios ataques por ransomware han puesto a prueba las infraestructuras informáticas de los servicios médicos de numerosos países. Precisamente por todo esto, las anomalías provocadas por ransomware fueron las seleccionadas para este entorno.

A partir de los requerimientos de los entornos elegidos, se decidió utilizar un enfoque basado en flujos, ya que era compatible con las restricciones impuestas: el uso de flujos reduce en órdenes de magnitud la cantidad de datos a examinar, permitiendo su uso con grandes volúmenes de tráfico de red y altas tasas de transferencia; los flujos no utilizan la carga útil de los paquetes, por tanto no se ven afectados por el hecho de que el tráfico de red esté cifrado. También se estudiaron los requerimientos computacionales de una variedad de métodos de aprendizaje automático, con el fin de encontrar el más adecuado para su evaluación en un entorno 5G según las restricciones en el tiempo de detección que dicho entorno impone por sus características. Se buscaron propuestas que admitieran estrictas restricciones en el tiempo de detección y capacidad de aprender patrones complejos de alto nivel. Adicionalmente, como parte del tercer objetivo a alcanzar y con el propósito de poder incorporar información contextual y temporal en la detección de anomalías, se estudiaron diversas opciones de vectores de características para seleccionar un conjunto de medidas que incorporaran esta información sin depender de la carga útil del paquete ni comprometer la privacidad del usuario. Entre las diferentes soluciones para la detección de

anomalías en tráfico de red en la literatura, se echó en falta un mayor número de propuestas que aprovecharan las ventajas del uso de NFV/SDN. Por ello, y siguiendo el cuarto objetivo propuesto, se consideró relevante plantear una arquitectura en el contexto de la detección de anomalías en el tráfico de red en entornos 5G que se basara en NFV/SDN, proporcionando los mecanismos necesarios para ajustar dinámicamente los recursos asignados a la detección continua de anomalías. Junto con el modelo y el vector de características, y como parte del quinto objetivo, se necesitó analizar los conjuntos de datos públicos para determinar al menos uno que contuviera tráfico real junto con tráfico generado por el tipo de anomalías escogidas, en este caso, botnets. El seleccionado fue el conjunto de datos denominado CTU [18]. Llegado a este punto, se realizó la selección de características, el aprendizaje del modelo con su ajuste de hiperparámetros y se evaluó el tiempo de detección usando los entornos de desarrollo para aprendizaje profundo más populares, cumpliendo el sexto y último objetivo. Los detalles relativos a este escenario 5G se encuentran en el Capítulo 1.

Tras esta primera propuesta, se decidió generalizar el concepto planteado en la red 5G al paradigma *Mobile Edge Computing*, adaptando la arquitectura para incluir una orquestación basada en políticas que permitieran la asignación inteligente de los recursos dedicados a nuestro sistema de detección de anomalías, con el fin de adaptarlo al volumen del tráfico soportado en los nodos próximos al usuario, es decir, los *Radio Access Network* (RAN) de la red 5G. Esta nueva propuesta se evaluó para determinar si realmente ofrecía una adaptación dinámica y en tiempo real a los cambios en las condiciones del tráfico de red. La descripción en detalle de lo anterior se puede consultar en el Capítulo 2.

Una vez evaluado el primer entorno, en el que la restricción principal es el número mínimo de detecciones por segundo que requería el volumen del tráfico soportado, se pasó a tratar el otro escenario, donde es crítico tanto el tiempo de detección como el de mitigación, al tiempo que el tráfico debía ser cifrado por privacidad. El escenario escogido como parte del segundo objetivo, fue el de los entornos clínicos integrados (ICE), que ofrecen la ventaja de poder ser virtualizados gracias a iniciativas como OpenICE [19]. En este segundo entorno se pretendía diseñar, implementar y validar un sistema automático e inteligente para detectar y mitigar ataques de ransomware que afecten a entornos ICE. La forma de detectar estos ataques serían las anomalías en el tráfico de la red del entorno clínico provocadas por la propagación del ransomware y el cifrado de carpetas compartidas. Precisamente por la necesidad de actuar rápidamente, la combinación NFV/SDN era especialmente adecuada al permitir aislar los dispositivos detectados como comprometidos gracias al paradigma SDN, al tiempo que NFV puede usarse para sustituir dinámicamente controladores software, modelos o incluso dispositivos virtualizados. Esto complementa el cuarto objetivo propuesto.

Al no encontrar un conjunto de datos adecuado en la literatura, se generaría un conjunto de datos etiquetados a partir de una configuración lo más realista posible de un entorno ICE basado en OpenICE y una selección de ransomware reciente. Puesto que el tráfico típico de los sensores ICE suele ser UDP y bastante regular en el tiempo, se incluirían equipos que accedan a bases de datos o carpetas compartidas en red, generando patrones de tráfico parecidos a los asociados al ransomware. Este conjunto de datos quedaría a disposición de la comunidad científica como parte del quinto objetivo propuesto. Como siguiente paso se planteó el diseño y validación de un modelo de aprendizaje automático que combinara un detector de anomalías supervisado y otro semi-supervisado para obtener lo mejor de ambos, y su integración en una arquitectura basada en NFV/SDN que proporcionara detección temprana y mitigación automática e inteligente. Para completar el sexto objetivo comprobando que efectivamente podría proteger ante la propagación de ransomware, se medirían tiempos de despliegue de dispositivos virtuales que reemplazaran a los comprometidos. Con este propósito se propuso la implementación de los dispositivos virtuales por medio de contenedores y máquinas virtuales corriendo en difer-

entes configuraciones hardware, con el fin de determinar también los requerimientos mínimos de hardware para conseguir la protección deseada. Los detalles relativos a este entorno pueden encontrarse en el Capítulo 3.

III Resultados

Esta tesis doctoral se presenta como compendio de publicaciones, por lo que los resultados están detallados en los artículos que la acompañan.

En el artículo que compone el Capítulo 1 se propone un sistema adaptativo basado en NFV/SDN para la detección de anomalías y ciberdefensa en el contexto de la arquitectura de comunicaciones móviles 5G. En este escenario se prevé que haya fluctuaciones inesperadas en el tráfico (por ejemplo, el tráfico en el RAN cercano a un estadio aumenta enormemente cuando hay un evento), en cuyo caso el sistema adapta sus recursos computacionales y sus elementos de inspección para optimizar y garantizar el proceso de detección. Esta adaptación se puede realizar fácil y naturalmente gracias al uso de funciones de red virtualizadas. Integrado en este sistema se incluye un modelo de detección de anomalías en dos niveles, donde el nivel inferior se ejecuta en el borde de la red y es un detector de síntomas de anomalías (anomalías locales al RAN en el que se ejecuta) que pueden formar parte de una anomalía más global que afecte a gran número de equipos finales. El nivel superior, por el contrario, se encuentra en la nube o un servidor central, recoge los síntomas convenientemente etiquetados provenientes de los niveles inferiores que se ejecutan en los diferentes RAN (opcionalmente se envía también información sobre los flujos culpables del síntoma por si se precisa un posterior procesamiento). Estos síntomas se ordenan por marca de tiempo con el fin de analizar la secuencia para intentar determinar la existencia de una anomalía de carácter global.

Debido a las peculiaridades del tráfico 5G y como fruto de la investigación descrita en el apartado de metodología, se decidió trabajar con vectores de características generados a partir de la secuencia de flujos de red recibidos (cada flujo representado en formato Netflow bidireccional) durante un cierto tiempo configurable. Estos vectores contienen datos agregados de todos los flujos recibidos en ese tiempo y, además, otros valores calculados únicamente a partir del último flujo recibido; siendo este último flujo el que determina la clase del vector de características resultante (normal/anómalo). El trabajar con flujos nos permite convertir el enorme volumen de datos de las redes 5G a un formato más manejable a costa de hacer más complejo si cabe el proceso de la detección de anomalías, ya que los patrones anómalos son más sutiles. Esto determinó la utilización de algoritmos de detección de anomalías basados en aprendizaje profundo por su habilidad para encontrar patrones complicados. Se analizaron los modelos de aprendizaje automático más populares en el contexto de detección de anomalías para determinar la escalabilidad a grandes volúmenes de datos, la estabilidad en el tiempo de predicción de de cada modelo, y el aprovechamiento del paralelismo proporcionado por hardware de propósito específico, como son las GPUs, entre otros factores. Se determinó que el mejor candidato para el primer nivel de nuestro sistema era una red neuronal profunda densa por ser un modelo que esencialmente se basa en productos de matrices, lo que permite un gran rendimiento al ejecutarse sobre una GPU, a la vez que su tiempo de predicción es independiente del proceso de entrenamiento. El segundo nivel recibe síntomas, y por tanto no tiene unas restricciones de tiempo de evaluación tan críticas. Debido a esto y a que debe de trabajar con secuencias de síntomas provenientes de diferentes RAN, interpretándolos como una secuencia de síntomas globales del sistema, se ha propuesto utilizar una red LSTM (*Long Short-Term Memory network*).

En este artículo se obtuvieron unos resultados preliminares del rendimiento en clasificación del primer nivel de nuestro detector de anomalías. La validación de este primer nivel del modelo

de aprendizaje automático situado en los RAN se realizó con el conjunto de datos de dominio público denominado CTU [18], que cumplía los requerimientos exigidos: provenir de tráfico real; ser suficientemente extenso como para obtener una muestra representativa del tráfico normal; tener la captura del tráfico tanto en formato pcap (sin la carga útil para evitar problemas de privacidad) como en el formato estándar de representación de flujos Netflow; y contener un número adecuado de familias de botnets con suficiente cantidad de tráfico anómalo. En este primer nivel aplicado a botnets conocidas se consiguió un valor de sensibilidad muy satisfactorio (0,9934), mientras que la precisión alcanzada quedó en únicamente un 0,8126. El ajuste de hiperparámetros durante el aprendizaje se hizo maximizando la sensibilidad con la intención de que el segundo nivel refinara el resultado. En el caso de la detección de botnets desconocidas (ninguna de ellas se usó en la etapa de entrenamiento), los resultados de este primer nivel en precisión estuvieron en el rango 0,40-0,95 (media de 0,686 y desviación típica de 0,1968) y los de sensibilidad en el rango 0,38-0,95 (media de 0,71 y desviación típica de 0,2421). Sin embargo en estas primeras pruebas, aparte de comprobar que el modelo detectaba adecuadamente, lo que se pretendía era demostrar la viabilidad del tiempo de ejecución del modelo propuesto en el primer nivel. Para ello se realizó una exhaustiva comparativa de rendimiento de las diferentes bibliotecas de desarrollo de modelos de aprendizaje profundo a la hora de predecir con redes neuronales densas. Los resultados mostraron que no había un único candidato vencedor, sino que la elección de la implementación estaría determinada por aspectos tales como el hardware disponible en un RAN o la tasa de recepción de flujos de red.

A partir del estudio realizado en el Capítulo 1, el Capítulo 2 extiende sustancialmente la versión preliminar de la arquitectura con nuevos módulos y funcionalidades para obtener una arquitectura enfocada a *Mobile Edge Computing* basada en NFV/SDN, que permite detectar anomalías de forma autónoma en tiempo real. Esta propuesta integra en el sistema el modelo de detección de anomalías de dos niveles basado en aprendizaje profundo planteado previamente, y propone el uso de políticas para conseguir un sistema de gestión dinámico y eficiente de los recursos computacionales usados en el proceso de detección. Gracias al uso de tres tipos de políticas y un proceso de orquestación a cargo de las acciones asociadas a dichas políticas, donde se integra el mecanismo de detección de anomalías, la arquitectura propuesta puede desplegar diversas acciones para asegurar una detección de anomalías efectiva en tiempo real. Como demostración de sus posibilidades se presentó un caso de uso basado en redes 5G, donde el método de detección de anomalías era el modelo de aprendizaje profundo en dos niveles descrito en el Capítulo 1. En este contexto se desarrollaron diferentes políticas dirigidas a mostrar cómo sería el proceso de adaptación dinámica del sistema propuesto, permitiendo a los RAN la asignación flexible y dinámica de recursos a la detección de los síntomas de anomalía al cambiar las condiciones del tráfico, por ejemplo. Esto se consigue gracias a las posibilidades que ofrecen las NFV de cambio en caliente de su configuración o el ajuste de los recursos del hardware virtualizado. Por ejemplo, cuando el tráfico es leve, se puede evaluar el modelo en una CPU, y cuando el tráfico es alto, se puede asignar hardware virtualizado adicional disponible, como puede ser una GPU, para que ejecute el modelo si la máquina virtual lo permite. Además, de esta manera, si el administrador recibe un nuevo modelo mejorado puede desplegarlo automáticamente.

En la experimentación se realizaron pruebas del rendimiento del modelo de detección de anomalías con el mismo conjunto de datos CTU para la detección de botnets. Para ello se reentrenó el detector de síntomas que se ejecuta en los RAN, implementado como una red profunda densa. En este reentrenamiento se empleó una exploración más amplia del espacio de hiperparámetros que la usada en el Capítulo 1, y se buscó maximizar la precisión. Finalmente se consiguió una precisión de 0,9537 y una sensibilidad de 0,9954 sobre botnets conocidas. Con respecto a botnets desconocidas, se obtuvo una precisión media de 0,8693 con desviación típica

de 0,1928 y una sensibilidad media de 0,4803 con desviación típica de 0.2353. Al ser las botnets una amenaza que permanece en el sistema durante largos periodos, es cuestión de tiempo que una baja sensibilidad termine detectándola, sin embargo, una baja precisión inundaría de falsos síntomas el segundo nivel. A continuación se tomaron los tiempos de ejecución del modelo estimados en el Capítulo 1, y se utilizaron para modelar una cota superior del tiempo de ejecución necesario para detectar una anomalía teniendo en cuenta que la evaluación del modelo debe hacerse usando lotes de vectores por razones de rendimiento. Esto se puso en práctica en un ejemplo donde se muestra la evolución del tráfico en el tiempo en una situación hipotética, y cómo el sistema se adaptaría, incluso anticipándose al punto crítico para realizar más suavemente la adaptación.

A continuación se describirán los principales resultados obtenidos en el Capítulo 3, donde se presenta la investigación enfocada a la detección de anomalías provocadas por ransomware en el contexto de los ICE. La aportación de esta parte del trabajo de investigación es la presentación de un sistema en tiempo real, automático e inteligente capaz de detectar, clasificar y mitigar ataques de ransomware en las habitaciones de hospital del futuro. La solución propuesta está integrada en la arquitectura ICE++ [20] y emplea técnicas de aprendizaje automático para detectar y clasificar la etapa de propagación de ataques de ransomware que afectan a ICE. Otra contribución relevante de este trabajo es el mecanismo propuesto de mitigación, que utiliza los paradigmas NFV/SDN para detener la propagación, aislando y reemplazando los dispositivos médicos infectados.

La propuesta que se presenta se basa en la creación, a partir de la agregación de secuencias de flujos en formato Netflow, de vectores de características que alimentan un módulo de análisis que contiene dos clasificadores, uno semi-supervisado utilizado para detectar anomalías más genéricas, y otro supervisado probabilista —cuya salida múltiple puede ser interpretada como una probabilidad de pertenencia a cada una de las clases— para clasificar ransomware. Un módulo de decisión y reacción, alimentado por un conjunto de reglas que actúan tanto en el plano de control como de datos, se encarga de tomar la salida de los dos clasificadores y reaccionar ante una detección de manera inteligente, por ejemplo, forzando una mitigación mediante el despliegue, gracias a la NFV, de un dispositivo virtual limpio si hay suficiente consenso entre los clasificadores, o alertando al administrador en los casos en que no haya consenso. Todo esto sucede en línea y en tiempo real, guardándose la información de los eventos en un log.

Estos vectores de características se utilizan para una fase fuera de línea que se encarga de la creación, etiquetado y particionado del conjunto de datos necesario para el entrenamiento de los modelos. A continuación se realiza una selección de características dependiente del modelo, y finalmente el entrenamiento, incluyendo el ajuste de hiperparámetros por validación cruzada para determinar el mejor modelo semi-supervisado y supervisado de una lista preseleccionada de métodos. Los modelos generados se integran en la etapa en línea descrita anteriormente

Para probar la efectividad de nuestra propuesta, se realizaron una serie de experimentos utilizando algunos de los malware más peligrosos y recientes: WannaCry, Petya, BadRabbit y PowerGhost. El primer paso fue recrear, usando OpenICE y máquinas virtuales, un entorno real clínico. Se decidió utilizar OpenICE por ser una implementación en codigo abierto bien conocida del estándar ICE, que permite simular dispositivos clínicos interoperables y controladores médicos capaces de desplegar escenarios en bucle cerrado. En este entorno virtual se adquirieron varias horas de tráfico de red normal. Después, para cada uno de los ransomware se partió de un entorno limpio, una de las máquinas fue contagiada con el ransomware, y se capturó el tráfico combinado resultante hasta el momento de en que se lograba la propagación del ransomware a otra máquina. A partir de este tráfico se generaron y etiquetaron los flujos y los vectores de características usados en la experimentación. Este conjunto de datos fue uno de los resultados de esta investigación y se encuentra a disposición de la comunidad científica.

Con respecto a los resultados en clasificación con este conjunto de datos, el método semi-supervisado que mejor rendimiento obtuvo en los experimentos fue OC-SVM (máquina de vector soporte monoclase), entrenada con tráfico normal, que obtuvo una precisión de 0,9232, una sensibilidad de 0,9997 y una tasa de falsos positivos de 0,046 a la hora de clasificar tráfico conjunto (normal y anómalo). En lo que respecta al método supervisado, cada uno de los métodos obtenía un rendimiento aceptable para cierta combinación de hiperparámetros. De entre dichos parámetros cabe destacar por su importancia el tamaño de la ventana de tiempo usada para la generación del vector de características, que terminó siendo fijada en 10 segundos. Se seleccionó Naive Bayes por ser el más sencillo de los tres evaluados y ofrecer un rendimiento similar para una ventana de 10 segundos, teniendo especialmente buen comportamiento en detectar ransomware desconocido. Los resultados muestran una precisión y sensibilidad del 99,99%. Es importante recordar que en las reglas que activan la mitigación se tienen en cuenta ambos predictores (OC-SVM y Naive Bayes) para tomar la decisión.

Por último se ha realizado un análisis del tiempo necesario para detectar y mitigar la propagación de ransomware en nuestra red, resultando que con la utilización de contenedores nuestro sistema podría recuperarse en un tiempo que va desde los 23,5 segundos en una Raspberry Pi 3, a 10,5 segundos en un ordenador personal. Estos tiempos son más que suficientes puesto que el tiempo de propagación del ransomware más bajo fue de 63 segundos. El desarrollo en detalle de estos resultados se encuentra en el Capítulo 3.

IV Conclusiones y Trabajo Futuro

Los ataques provocados por botnets y ransomware suponen un alto porcentaje de las amenazas más peligrosas en ciberseguridad, provocando incalculables pérdidas cada año. Las redes que nos trae el futuro próximo son un reto a la hora de aplicar métodos de detección de estas amenazas en el tráfico de red basados en inspección profunda de paquetes, debido principalmente a las restricciones que imponen aspectos tales como las altas velocidades de transmisión, enormes volúmenes de datos, la incesante expansión del Internet de las Cosas y la cada vez mayor adopción del cifrado el tráfico de red. A todo esto hay que añadir la facilidad para generar nuevas versiones modificadas de malware con el fin de evitar la detección.

En esta tesis se proponen tres soluciones basadas en NFV/SDN y aprendizaje automático para la detección de anomalías en el tráfico de red en tres escenarios representativos del futuro próximo:

- Un sistema adaptativo para la detección de anomalías en red en el contexto de las comunicaciones móviles 5G que permite optimizar sus recursos computacionales garantizando poder ejecutar el proceso de detección.
- Una extensión del anterior para derivar una arquitectura enfocada a *Mobile Edge Computing* que incorpora políticas que permiten desplegar acciones que garanticen la detección en tiempo real y efectiva de las anomalías.
- Un sistema en tiempo real, automático e inteligente capaz de detectar, clasificar y mitigar ataques de ransomware en los entornos clínicos integrados de las habitaciones de hospital del futuro.

El nexo de unión de estas tres soluciones ha sido la utilización de un enfoque novedoso para la detección de anomalías en el tráfico de las redes de datos. Dicho enfoque está basado en la idea de que en una secuencia de flujos de red todos los flujos previos al último recibido se pueden interpretar como un contexto de ese último flujo. A partir de una secuencia de flujos se genera

un vector de características agregadas, con el propósito de condensar información contextual en un solo vector al que aplicar técnicas de detección de anomalías puntuales. Esta forma obtener un vector que condensa información contextual del tráfico de red para la detección de anomalías, se ha integrado en las tres soluciones propuestas y validado en los tres escenarios indicados.

Entre las ventajas de la utilización de flujos están las siguientes: permite no depender del acceso a la carga útil del paquete, por lo que funciona con tráfico cifrado; reduce en órdenes de magnitud la cantidad de datos a analizar; y facilita la adquisición de conjuntos de datos de entrenamiento al ser compatible con la privacidad del usuario. Sin embargo, también por el hecho de no utilizar la carga útil del paquete, los patrones a identificar en los vectores resultantes son más complejos —por lo que no se prestan fácilmente al uso de reglas—, a lo que hay que sumar la necesidad de que haya una cierta protección ante variaciones desconocidas. Los resultados mostrados apoyan la hipótesis de que los métodos de aprendizaje automático, con la suficiente cantidad de datos, pueden llegar a identificar dichos patrones y generalizarlos a nuevas versiones del malware, convirtiéndose en una valiosa herramienta en un futuro próximo.

Escenarios

En el escenario 5G y gracias a la flexibilidad que su estándar ofrece, se ha propuesto una arquitectura basada en NFV/SDN en la que se integra un mecanismo de detección de anomalías a dos niveles. El nivel inferior, que se ejecuta en cada RAN, realiza una detección de síntomas de anomalías locales, mientras que el nivel superior, centralizado, recibe la secuencia de síntomas de anomalías de todos los detectores en los diferentes RAN, convenientemente etiquetadas y acompañadas de su contexto, y las convierte en una secuencia temporal de síntomas que se analiza en busca de anomalías globales. Las restricciones de tiempo que imponen las altas velocidades de transferencia y el volumen de datos que se espera en 5G llevaron a la decisión de utilizar redes neuronales profundas para la detección de los síntomas, tanto por tener un tiempo de ejecución en predicción estable e independiente del entrenamiento, como por su eficiencia al ejecutarse sobre GPUs. Se realizó un estudio de los diferentes entornos de aprendizaje profundo y se concluyó que los recursos hardware disponibles (por ejemplo, si se debe usar CPU o hay una GPU o varias) determinan la implementación del modelo de aprendizaje usado para la detección. El primer nivel de este modelo, situado en los RAN de la infraestructura 5G, fue validado utilizando el conjunto de datos público CTU, que contiene tráfico real capturado en una universidad, tanto limpio como incluyendo tráfico de varias familias de botnets. Se concluyó que el modelo elegido junto con la utilización de agregaciones de flujos, era una combinación que permitía obtener un rendimiento en clasificación adecuado (puesto que el segundo nivel se encargaría de la decisión final) como, más importante, permitir la evaluación del modelo en un tiempo acotado aceptable para el sistema.

En el escenario de *Mobile Edge Computing*, se extendió sustancialmente la anterior solución para incluir políticas y un proceso de orquestación a cargo de acciones asociadas a dichas políticas, que permiten obtener un sistema de gestión dinámico y eficiente de los recursos computacionales usados en el proceso de detección en tiempo real de anomalías. Se incluyó un estudio, en el contexto de redes 5G, de los tiempos de ejecución necesarios para la detección de las anomalías y el comportamiento dinámico del sistema en un caso de uso de adaptación a la evolución del tráfico.

Con respecto al estudio del rendimiento del modelo de dos niveles en conjunto, se estaban manteniendo conversaciones con una compañía de telecomunicaciones para realizar una adquisición de datos. Desgraciadamente no se consiguió un acuerdo, y por tanto la captura de un conjunto de datos reales y etiquetados proveniente de varios RAN que permita validar el segundo nivel del sistema queda como trabajo futuro.

En el último escenario elegido, los entornos clínicos integrados del futuro, una contribución es la incorporación de un mecanismo de detección de anomalías, provocadas en el tráfico de red por la propagación de ransomware, a la arquitectura ICE++ basada en NFV/SDN. Precisamente el uso de NFV/SDN permite otra de las contribuciones de esta propuesta, que es la detección y mitigación inteligente y dinámica de dicho ransomware. Para la detección se ha propuesto la utilización de un método de detección semi-supervisado y otro supervisado para unificar sus predicciones mediante un mecanismo de reglas que proporciona una gestión automática e inteligente tanto de la mitigación en ciertos casos, como de las alertas al administrador en otros. La modificación de la arquitectura ICE++ incluyó tanto la operativa fuera de línea —necesaria para la adquisición, limpieza y etiquetado de los conjuntos de datos, así como la selección de características y entrenamiento de los modelos— como la operativa en línea, que se encarga de la generación de los vectores de características en tiempo real y su clasificación en anómalos o normales en el caso del algoritmo semi-supervisado, y de proporcionar una probabilidad de pertenencia a cada clase de ransomware en el caso del supervisado. A falta de un conjunto de datos disponible públicamente, se creó un escenario clínico realista virtualizado con OpenICE del que se capturó tráfico limpio y tráfico anómalo provocado por la propagación del ransomware. Haber puesto a disposición de la comunidad ese conjunto de datos es otra de las contribuciones de esta tesis. El conjunto de datos se usó para validar la capacidad de detección de los modelos y realizar un estudio de la velocidad de reacción de la arquitectura en detección/mitigación sobre diferente hardware. El sistema se mostró capaz de detectar y mitigar tanto ransomware conocido como nuevo en un tiempo sustancialmente menor que el necesario para que dicho ransomware se propague y con una gran efectividad.

Trabajo futuro

Una de las líneas de trabajo futuro de esta tesis, consiste en la aplicación de modelos de aprendizaje profundo más complejos como son redes convolucionales y los recientemente propuestos *transformers* basados en mecanismos de atención. La utilización de secuencias de flujos se presta a su análisis por estos métodos que han demostrado su buen desempeño en tareas de reconocimiento del lenguaje natural, entre otras, mejorando sustancialmente los resultados conseguidos con redes recurrentes en la identificación de patrones complejos.

Otro aspecto que es de suma importancia para poder investigar en la dirección que esta tesis plantea, es la existencia de conjuntos de datos realistas y suficientemente variados para poder aplicar técnicas de aprendizaje profundo como las mencionadas. Actualmente el grupo de investigación al que pertenezco sigue intentando llegar a un acuerdo con alguna compañía de comunicaciones móviles para conseguir tráfico real anonimizado.

Por último, otra futura línea de trabajo consiste en aplicar las técnicas aquí desarrolladas a otros entornos, como puede ser el de redes industriales, que tradicionalmente han estado aisladas y, precisamente por eso, presentan vulnerabilidades que quedan expuestas con la gradual apertura de estos sistemas a Internet. La utilización de la red para controlar procesos industriales abre la puerta a todo tipo de sabotajes por malware que altere un tráfico que en estos entornos suele tener patrones muy estables de funcionamiento. Estos patrones precisamente hacen que la detección de anomalías sea una opción a tener en cuenta. Ya se ha contactado con empresas industriales de la región que han considerado seriamente la posibilidad de ofrecer sus instalaciones para la captura de tráfico y validación de las soluciones planteadas.

Publicaciones que componen

la tesis doctoral

1

A self-adaptive deep learning-based system for anomaly detection in 5G networks

Título:	A Self-Adaptive Deep Learning-Based System for Anomaly Detection in 5G Networks
Autores:	Lorenzo Fernández Maimó, Ángel Luis Perales Gómez, Félix J. García Clemente, Manuel Gil Pérez, Gregorio Martínez Pérez
Revista:	IEEE Access
Factor de Impacto:	3,557 JCR Q1 (2017)
Editor:	IEEE
Volumen:	6
Páginas:	7700–7712
Año:	2018
Mes:	Febrero
DOI:	http://dx.doi.org/10.1109/ACCESS.2018.2803446
Estado:	Publicado

Abstract

The upcoming fifth-generation (5G) mobile technology, which includes advanced communication features, is posing new challenges on cybersecurity defense systems. Although innovative approaches have evolved in the last few years, 5G will make existing intrusion detection and defense procedures become obsolete, in case they are not adapted accordingly. In this sense, this paper proposes a novel 5G-oriented cyberdefense architecture to identify cyberthreats in 5G mobile networks efficient and quickly enough. For this, our architecture uses deep learning techniques to analyze network traffic by extracting features from network flows. Moreover, our proposal allows adapting, automatically, the configuration of the cyberdefense architecture in order to manage traffic fluctuation, aiming both to optimize the computing resources needed in each particular moment and to fine tune the behavior and the performance of analysis and detection processes. Experiments using a well-known botnet data set depict how a neural network model reaches a sufficient classification accuracy in our anomaly detection system. Extended experiments using diverse deep learning solutions analyze and determine their suitability and performance for different network traffic loads. The experimental results show how our architecture can self-adapt the anomaly detection system based on the volume of network flows gathered from 5G subscribers' user equipments in real-time and optimizing the resource consumption.

2

Dynamic management of a deep learning-based anomaly detection system for 5G networks

Título:	Dynamic management of a deep learning-based anomaly detection system for 5G networks
Autores:	Lorenzo Fernández Maimó, Alberto Huertas Celdrán, Manuel Gil Pérez, Félix J. García Clemente, Gregorio Martínez Pérez.
Revista:	Journal of Ambient Intelligence and Humanized Computing
Factor de impacto:	1,423 JCR Q3 (2017)
Editor:	Springer Berlin Heidelberg
Año:	2018
Mes:	Mayo
DOI:	http://dx.doi.org/10.1007/s12652-018-0813-4
Estado:	Publicado en línea

Abstract

Fog and mobile edge computing (MEC) will play a key role in the upcoming fifth generation (5G) mobile networks to sup port decentralized applications, data analytics and management into the network itself by using a highly distributed compute model. Furthermore, increasing attention is paid to providing user-centric cybersecurity solutions, which particularly require collecting, processing and analyzing significantly large amount of data traffic and huge number of network connections in 5G networks. In this regard, this paper proposes a MEC-oriented solution in 5G mobile networks to detect network anomalies in real-time and in autonomic way. Our proposal uses deep learning techniques to analyze network flows and to detect network anomalies. Moreover, it uses policies in order to provide an efficient and dynamic management system of the computing resources used in the anomaly detection process. The paper presents relevant aspects of the deployment of the proposal and experimental results to show its performance.

3

Intelligent and dynamic ransomware spread detection and mitigation in integrated clinical environments

Título:	Intelligent and Dynamic Ransomware Spread Detection and Mitigation in Integrated Clinical Environments
Autores:	Lorenzo Fernández Maimó, Alberto Huertas Celdrán, Ángel Luis Perales Gómez, Félix J. García Clemente, James Weimer, Insup Lee.
Revista:	Sensors
Índice de impacto:	2,475 JCR Q2 (2017)
Editor:	Multidisciplinary Digital Publishing Institute
Volumen:	19
Número:	5
Páginas:	1114
Año:	2019
Mes:	Enero
DOI:	http://dx.doi.org/10.3390/s19051114
Estado:	Publicado

Abstract

Medical Cyber-Physical Systems (MCPS) hold the promise of reducing human errors and optimizing healthcare by delivering new ways to monitor, diagnose and treat patients through integrated clinical environments (ICE). Despite the benefits provided by MCPS, many of the ICE medical devices have not been designed to satisfy cybersecurity requirements and, consequently, are vulnerable to recent attacks. Nowadays, ransomware attacks account for 85% of all malware in healthcare, and more than 70% of attacks confirmed data disclosure. With the goal of improving this situation, the main contribution of this paper is an automatic, intelligent and real-time system to detect, classify, and mitigate ransomware in ICE. The proposed solution is fully integrated with the ICE++ architecture, our previous work, and makes use of Machine Learning (ML) techniques to detect and classify the spreading phase of ransomware attacks affecting ICE. Additionally, Network Function Virtualization (NFV) and Software Defined Networking (SDN)paradigms are considered to mitigate the ransomware spreading by isolating and replacing infected devices. Different experiments returned a precision/recall of 92.32%/99.97% in anomaly detection, an accuracy of 99.99% in ransomware classification, and promising detection and mitigation times. Finally, different labelled ransomware datasets in ICE have been created and made publicly available.

Referencias

[1] Nokia. Nokia threat intelligence report. https://pages.nokia.com/T003B6-Threat-Intelligence-Report-2019.html, 2019. Accedido el 12 de abril de 2019.

[2] V. Richariya, U. P. Singh, and R. Mishra. Distributed approach of intrusion detection system: Survey. *International Journal of Advanced Computer Research*, 2(6):358–363, 2012.

[3] J. Yu, B. Yang, R. Sun, and Y. Chen. FPGA-based parallel pattern matching algorithm for network intrusion detection system. In *2009 International Conference on Multimedia Information Networking and Security*, pages 458–461, November 2009.

[4] Y.-M. Hsiao, M.-J. Chen, Y.-S. Chu, and C.-H. Huang. High-throughput intrusion detection system with parallel pattern matching. *IEICE Electronics Express*, 9(18):1467–1472, September 2012.

[5] A. L. Buczak and E. Guven. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2):1153–1176, October 2016.

[6] Q. A. Tran, F. Jiang, and J. Hu. A real-time NetFlow-based intrusion detection system with improved BBNN and high-frequency field programmable gate arrays. In *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 201–208, June 2012.

[7] J. Gardiner and S. Nagaraja. On the security of machine learning in malware C&C detection: A survey. *ACM Computing Surveys*, 49(3):59:1–59:39, December 2016.

[8] Cisco. Encrypted traffic analytics (white paper). https://www.cisco.com/c/dam/en/us/solutions/collateral/enterprise-networks/enterprise-network-security/nb-09-encrytd-traf-anlytcs-wp-cte-en.pdf, 2019. Accedido el 12 de abril de 2019.

[9] Gillian Mohney. Hospitals remain key targets as ransomware attacks expected to increase. https://abcnews.go.com/Health/hospitals-remain-key-targets-ransomware-attacks-expected-increase/story?id=47416989, 2017. Accedido el 4 de marzo de 2019.

[10] Tiina Kovanen, Gil David, and Timo Hämäläinen. Survey: Intrusion detection systems in encrypted traffic. In *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, pages 281–293. Springer, 2016.

[11] Antonio Pastor, Alberto Mozo, Diego R Lopez, Jesus Folgueira, and Angeliki Kapodistria. The mouseworld, a security traffic analysis lab based on nfv/sdn. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, page 57. ACM, 2018.

[12] Wang, W and Sheng, Y and Wang, J et al. HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. *IEEE Access*, 6:1792–1806, 2018.

[13] Palo Alto Networks Threat Intelligence Team (Unit 42). Unit 42 Finds New Mirai and Gafgyt IoT/Linux Botnet Campaigns. https://unit42.paloaltonetworks.com/unit42-finds-new-mirai-gafgyt-iotlinux-botnet-campaigns/, 2018. Accedido el 19 de abril de 2019.

[14] Heimdal Security. The Top Ten: Most Dangerous Malware That Can Empty Your Bank Account. https://heimdalsecurity.com/blog/top-financial-malware/, 2016. Accedido el 19 de abril de 2019.

[15] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

[16] K. Cabaj and W. Mazurczyk. Using software-defined networking for ransomware mitigation: The case of cryptowall. *IEEE Network*, 30(6):14–20, November 2016.

[17] Greg Cusack, Oliver Michel, and Eric Keller. Machine learning-based detection of ransomware using sdn. In *Proceedings of the 2018 ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization*, SDN-NFV Sec'18, pages 1–6, New York, NY, USA, 2018. ACM.

[18] S. Garcia, M. Grill, J. Stiborek, and A. Zunino. An empirical comparison of botnet detection methods. *Computers & Security*, 45:100–123, September 2014.

[19] Jeffrey Plourde Arney, David and Julian M. Goldman. Openice medical device interoperability platform overview and requirement analysis. In *Biomedical Engineering / Biomedizinische Technik*, pages 39–47, November 2017.

[20] A. Huertas, F. J. Garcia, J. Weimer, and I. Lee. Ice++: Improving security, qos, and high availability of medical cyber-physical systems through mobile edge computing. In *Proceedings of the IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–8, Sept 2018.