Universitat Oberta de Catalunya uoc.edu

Desplegar la herramienta "Zeek" y su posterior explotación para el análisis de actividades sospechosas en la red

Nombre Estudiante: Sergio Álvarez Rubio
Máster en Seguridad de las Tecnologías de la Información y las Comunicaciones
Trabajo de Fin de Máster – Área Hacking

Nombre Consultor/a: Borja Guaita Pérez
Nombre Profesor/a responsable de la asignatura: Víctor García Font

31/12/2019

Agradecimientos:

Quiero agradecer en primer lugar a mi gran amor y compañera de viaje, Inma, ya que con su amor, paciencia, ánimos e insistencia en realizar este Máster me ha permitido llegar a cumplir hoy un sueño más.

A mis amores, Rubén y Víctor, que, aunque siempre con bromas sobre la edad en la que afrontaba estos estudios, me miraban sorprendidos, pero a la vez orgullosos, cuando salía publicada una nueva nota de las entregas parciales de todas las asignaturas. Me han dado la fuerza de voluntad para proseguir hasta el final.

Agradecer también a mis padres, mis hermanas y, sin olvidarme a los que ya no están junto a nosotros, mis abuelos porque gran parte de lo que soy se lo debo a ellos.

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Despliegue de la herramienta "Zeek" y su posterior explotación para el análisis de actividades sospechosas en la red*
Nombre del autor:	*Sergio Álvarez Rubio*
Nombre del consultor/a:	*Borja Guaita Pérez*
Nombre del PRA:	*Víctor García Font*
Fecha de entrega:	12/2019
Titulación:	*Máster en Seguridad de las Tecnologías de la Información y las Comunicaciones*
Área del Trabajo Final:	*Hacking*
Idioma del trabajo:	*Español*
Palabras clave:	*Network Intrusion Detection System, Seguridad del Paciente, Zeek-ELK Stack*

Resumen del Trabajo:

La finalidad principal de este trabajo es valorar qué ventajas nos ofrece la implementación de un Sistema de Detección de Intrusiones en combinación con herramientas de transformación de logs para su almacenaje en Bases de Datos NoSQL y su posterior explotación mediante el uso de herramientas avanzadas de creación de distintas gráficas que nos ayuden a la detección de comportamientos anómalos en nuestra red dentro del ámbito laboral y a coste reducido, aprovechando los recursos disponibles.

La particularidad de nuestro entorno laboral hace que estén conectados a la red muchos equipos electromédicos con Windows embebido y, dada la dificultad de aplicación de parches de seguridad, suponen una importante amenaza para la seguridad de la información y, más concretamente, para la seguridad del paciente.

Para ello, se ha utilizado Zeek como Sistema de Detección de Intrusiones en Red (NIDS) y ELK Stack para la transformación de logs (Beats), almacén de datos (Elastic Search) y para la posterior explotación y visualización gráfica de la información, Machine Learning y SIEM (Kibana).

Se ha utilizado hardware virtualizado para ELK Stack en Windows Server 2012 R2 y PCs, cuyo valor residual es 0, para instalar Zeek en Linux como Sniffer de Red utilizando configuración de puertos espejo en switch gestionado.

Se ha podido visualizar varias anomalías, con lo que no ha sido necesaria la simulación de ataques o infecciones, que nos ha llevado a la detección de dos infecciones por Wannacry. Concluyendo que el entorno implementado nos ayuda a la detección de anomalías en red.

Abstract:

The main purpose of this work is to assess the advantages offered by the implementation of an Intrusion Detection System in combination with log transformation tools for storage in NoSQL Databases and its subsequent exploitation through the use of advanced tools for creating different graphs that help us to detect anomalous behaviors in our network within the workplace and at reduced cost, taking advantage of the available resources.

The particularity of our work environment means that many electromedical devices with embedded Windows are connected to the network and, given the difficulty of applying security patches, pose a significant threat to information security and, more specifically, to the security of the patient.

For this, Zeek has been used as a Network Intrusion Detection System (NIDS) and ELK Stack for the transformation of logs (Beats), data warehouse (Elastic Search) and for the subsequent exploitation and graphic visualization of information, Machine Learning and SIEM (Kibana).

Virtualized hardware has been used for ELK Stack on Windows Server 2012 R2 and PCs, whose residual value is 0, to install Zeek on Linux as a Network Sniffer using mirror port configuration on managed switch.

It has been possible to visualize several anomalies, which has not been necessary to simulate attacks or infections, which has led us to the detection of two Wannacry infections. Concluding that the implemented environment helps us to detect network anomalies.

Índice

1. Introducción 1

1.1. Contexto y justificación del Trabajo 1

¿Por qué un IDS? 1

¿Qué es un SIEM? 2

1.2. Objetivos del Trabajo 3

2. Estado del arte 4

3. Metodología y Planificación 6

3.1. Enfoque y método seguido 6

3.2. Lista de Tareas 6

Plan de Trabajo 6

Recopilación de Información, Análisis y Planteamiento 7

Implementación 8

Memoria Final 9

3.3. Identificación de Riesgos 9

3.4. Planificación del Trabajo 11

4. Recopilación de información, análisis y planteamiento 13

4.1. Herramientas IDS 13

Definición de IDS 13

IDS en la actualidad 15

4.2. Zeek 16

4.3. Bases Datos NoSQL 17

Adaptabilidad JSON 18

Comparación en términos de excavación de datos 19

Problema de pérdida de datos en ElasticSearch 19

Búsqueda de datos 19

Decisión 20

4.4. SIEMs 20

4.5. Plataforma ELK 21

Ventajas de la plataforma ELK 22

BEATS 22

Kibana 23

Detección de Anomalías con Machine Learning 25

SIEM Elastic 27

Otros componentes ELK Stack *30*
4.6. Seguridad del Paciente asociada al uso de las TIC **31**
Concepto de Seguridad del Paciente *31*
Beneficios del uso de las TIC en la Seguridad del Paciente *31*
Riesgos del uso de las TIC en la Seguridad del Paciente *33*
5. Implementación **36**
5.1. Aprovisionamiento de recursos hardware y software **36**
5.2. Instalación y configuración ELK Stack **37**
Instalación de Elasticsearch *37*
Instalación de Kibana *38*
5.3. Instalación y configuración ZEEK (IDS BRO) **39**
5.4. Integración Zeek-ELK Stack **41**
5.5. Integración SIEM **47**
6. Producto final **49**
6.1. Test y Puesta en Marcha **49**
Comenzamos con la ayuda del Machine Learnig *50*
Investigación equipos infectados por Wannacry *54*
Ramsomware Bluekeep *58*
Caída de servidor Oracle *60*
6.2. Otras configuraciones **62**
Definición de espacios de trabajo *62*
Definición de Dashboards *63*
Trabajos de Rollups *64*
7. Conclusiones **65**
7.1. Consecución de objetivos marcados **65**
7.2. Conclusiones del trabajo **66**
7.3. Líneas de trabajo futuro pendientes **67**
8. Glosario **69**
9. Bibliografía **70**
9.1. Referencias bibliográficas: **70**
9.2. Referencias Web: **70**
9.3. Vídeos Visitados: **71**
10. Anexo I - Características ELK Stack Basic **72**
11. Anexo II – Logs generados por Zeek **75**

Lista de figuras

Figura 1- Clasificación de los Sistemas de Detección de Intrusos 4
Figura 2- Comparativa IDS 2019 15
Figura 3 - Componentes necesarios para SIEM (Fuente Elastic) 28
Figura 4 - SIEM Vista Hosts 29
Figura 5 - SIEM Vista Network 29
Figura 6 - SIEM Timeline 30
Figura 7 - Almacenamiento ELK Stack 36
Figura 8 - ElasticSearch como Servicio 37
Figura 9 - Instalación Elastic Search Licencia 38
Figura 10 - Instalación Elastic Search Licencia 38
Figura 11 - NIDS en línea 39
Figura 12 - NIDS con Port Mirroring 39
Figura 13 - Logs tráfico de red detectado por Zeek 41
Figura 14 - Integración Zeek & ELK Stack 41
Figura 15- Ubicación Logs Zeek para Filebeat 43
Figura 16 - Transformación Logs Filebeat 43
Figura 17 - Módulos Integración Zeek-Filebeat GitHub 44
Figura 18 - Creación de Índice Paso 1 44
Figura 19 - Creación de Índice Paso 2 45
Figura 20 - Nodo Centro 1 “HPONZEEK1” 45
Figura 21 - Nodo Centro 2 “HGUAZEEK1” 46
Figura 22 - Objetos creados por defecto para Zeek 46
Figura 23 – Objetos creados por defecto para Apache 47
Figura 24 - Acceso módulo SIEM 47
Figura 25 - Eventos Filebeat Zeek SIEM 48
Figura 26 - Acceso Machine Learning ELK Stack 50
Figura 27 - Selección de Índice para Machine Learning 50
Figura 28 - Métricas "source" Machine Learning 51
Figura 29 - Campos "source" Machine Learning 52
Figura 30 - Métricas "destination" Machine Learning 53
Figura 31 - Campos "destination" Machine Learning 53
Figura 32 - Log Antivirus Corporativo 54
Figura 33 - Top Ten conexiones puerto 445 (SMB) 55
Figura 34 - Mapa Mundial intentos infección Wannacry 56
Figura 35 – Campaña en Twitter de Microsoft sobre Bluekeep 58
Figura 36 - Posible infección por Bluekeep 59
Figura 37 - Conexiones excesivas a Oracle ocasionando reinicio del servidor 61
Figura 38 - Ejemplo Espacio HPONIENTE 62
Figura 39 – Ejemplo Dashboard definido 63
Figura 40 - Trabajos mantenimiento de Índices 64

1. Introducción

1.1. Contexto y justificación del Trabajo

Si hace unos años todas las miradas e inversiones estaban dirigidas hacia la alta disponibilidad con sistemas redundantes, clústeres, sistemas de copias de seguridad, etc., hoy en día la seguridad se está convirtiendo en una obsesión para muchas empresas debido al crecimiento constante de las ciberamenazas.

De forma general, se entiende que mantener un sistema seguro, o fiable, consiste básicamente en garantizar cuatro aspectos: confidencialidad, integridad, disponibilidad y no repudio.

- La confidencialidad, requiere que la información sea accesible únicamente por aquellos que estén autorizados.

- La integridad, que la información se mantenga inalterada ante accidentes o intentos maliciosos.

- La disponibilidad significa que el sistema informático se mantenga trabajando sin sufrir ninguna degradación en cuanto a accesos y ofrezca los recursos que requieran los usuarios autorizados cuando éstos los necesiten.

- Por último, el no repudio garantiza al emisor que la información fue entregada y ofrece una prueba al receptor del origen de la información recibida.

Mantener los recursos de información y telecomunicaciones protegidos contra infecciones, amenazas, sabotajes, extraños y/ intrusos, es una de las principales prioridades para cualquier organización. Para ello, nos hemos familiarizado con algunas herramientas y/o conceptos más comúnmente utilizados como Antivirus, Anti-malware, Firewalls, Parches de seguridad, etc.

Pero también existen otras herramientas utilizadas más en el ámbito de medianas-grandes organizaciones para detección de posibles amenazas, intrusos o comportamientos anómalos que pueden escapar a las herramientas anteriores. Para ello, en las mencionadas organizaciones, se ha extendido el uso de la combinación de IDS (Intrusion Detection System) y SIEM (Security Information & Event Management).

¿Por qué un IDS?

"La mejor manera de entender la necesidad de incorporar estos elementos podría ser la comparación entre la seguridad de una red informática y la seguridad de un edificio: las puertas de entrada ejercen un primer nivel de control de acceso, pero normalmente no nos quedamos aquí; instalaremos detectores de movimiento o cámaras de vigilancia en puntos clave del edificio para detectar la existencia de personas no autorizadas, o que hacen un mal uso de los recursos, poniendo en peligro la seguridad. Además, existirán vigilantes de seguridad, libros de registro en los que se apuntará a todo el

personal que accede a un determinado departamento que consideramos crítico, etcétera. Toda esta información se procesa desde una oficina de control de seguridad donde se supervisa el registro de las cámaras y se llevan los libros de registro. Todos estos elementos, proyectados en el mundo digital, configuran lo que se conoce en el ámbito de la seguridad de redes informáticas como mecanismos de detección.". Módulo 2-Sistemas de Detección de Intrusos en la red de la asignatura Seguridad de Redes de la UOC.

¿Qué es un SIEM?

Con la utilización de un IDS detectamos la necesidad de funcionalidades adicionales para completar el proceso de detección. Para ello, necesitamos el uso de nuevos elementos encargados de gestionar sistemas y técnicas de detección heterogéneas, con el objetivo final de poder controlar y reaccionar apropiadamente ante escenarios de intrusión más complejos. Estos nuevos elementos, generalizados con el nombre de SIEM, tienen por objetivo facilitar la gestión de grandes volúmenes de información generados por herramientas de detección, no únicamente IDS, sino también eventos generados a partir de sistemas de cortafuegos, escáneres de vulnerabilidades o incluso sistemas antivirus. Dichos elementos proporcionarán técnicas necesarias para normalizar los eventos recibidos y, finalmente, realizar tareas de fusión de información y correlación de alertas.

En la elección de este TFM ha influido la necesidad de implementación de un sistema de detección de actividades sospechosas en el Hospital en el que trabajo y en la que existen multitud de equipos diferentes interconectados como:

- Equipos electromédicos con sistemas operativos embebidos
- Servidores de Bases de Datos Oracle con AIX y Linux
- Servidores de Bases de Datos MySql
- Servidores DHCP
- Servicios de Active Directory
- Servidores Web
- Dispositivos móviles
- Equipos PCs

El hospital se encuentra, en principio, "aislado" y su acceso exterior es exclusivo a través de una VPN. Aunque sí que se conectan dispositivos móviles a la red, pendrives y está permitida la navegación a internet sólo a los puertos 80 y 443. Además, se dispone de Firewall de salida a la VPN y antivirus empresarial.

De esta forma, con este TFM se pretende detectar comportamientos anómalos o sospechosos que puedan afectar a la Seguridad de la Información pero se hará especial énfasis en aquellos que afecten o puedan afectar a la Seguridad del Paciente derivado

del uso de las nuevas tecnologías de la información y, en tal caso, notificarlo al Observatorio para la Seguridad del Paciente.

Las principales dificultades se presentarán para poder definir qué eventos, comportamientos o incidentes pueden afectar de algún modo a la Seguridad del Paciente pero es un reto muy interesante y que, en cualquier caso, puede salvar vidas.

1.2. Objetivos del Trabajo

El objetivo principal de este trabajo es la implantación de un sistema completo de análisis y detección de actividades sospechosas y, con especial atención, a aquellas que puedan poner en peligro la seguridad del paciente debido al uso de las nuevas tecnologías de la información.

Para ello es necesario alcanzar otros objetivos previos en los que se apoyará:

- ✓ Identificar y describir cómo puede afectar una intrusión en un equipo electromédico en la seguridad del paciente.
- ✓ Identificar y describir cómo puede afectar una intrusión en la base de datos de historia clínica en la seguridad del paciente.
- ✓ Analizar y comprender cómo puede ayudar un sistema de detección de actividades sospechosas en la identificación de las amenazas para el problema que se nos plantea en este trabajo.
- ✓ Analizar y comprender cómo el Machine Learning puede ayudar en la detección de actividades sospechosas y su aplicación en nuestro sistema a implementar.
- ✓ Estudiar y aplicar sistemas de almacenamiento de gran volumen de datos para la ayuda en la identificación de actividades sospechosas en la red.
- ✓ Estudiar y decidir qué sistema de detección de intrusiones es el adecuado a la problemática planteada y qué herramientas adicionales son necesarias para obtener el resultado deseado.
- ✓ Investigar distintas fuentes externas de listas de reputación que puedan ayudar en la detección de actividades sospechosas y su aplicación en la implementación de nuestro sistema.
- ✓ Exponer las conclusiones del trabajo realizado.

2. Estado del arte

Voy a intentar avanzar algo de información de en qué punto nos encontramos actualmente con respecto al trabajo que se quiere realizar, aunque a lo largo del desarrollo del proyecto iremos avanzando y profundizando en ello.

Tras realizar búsquedas sobre trabajos de fin de máster, grado y, principalmente, en Internet, he encontrado que existe bastante información acerca de la implantación de distintos IDS y sobre los distintos tipos de IDS que existen y que se puede ver en el siguiente esquema:

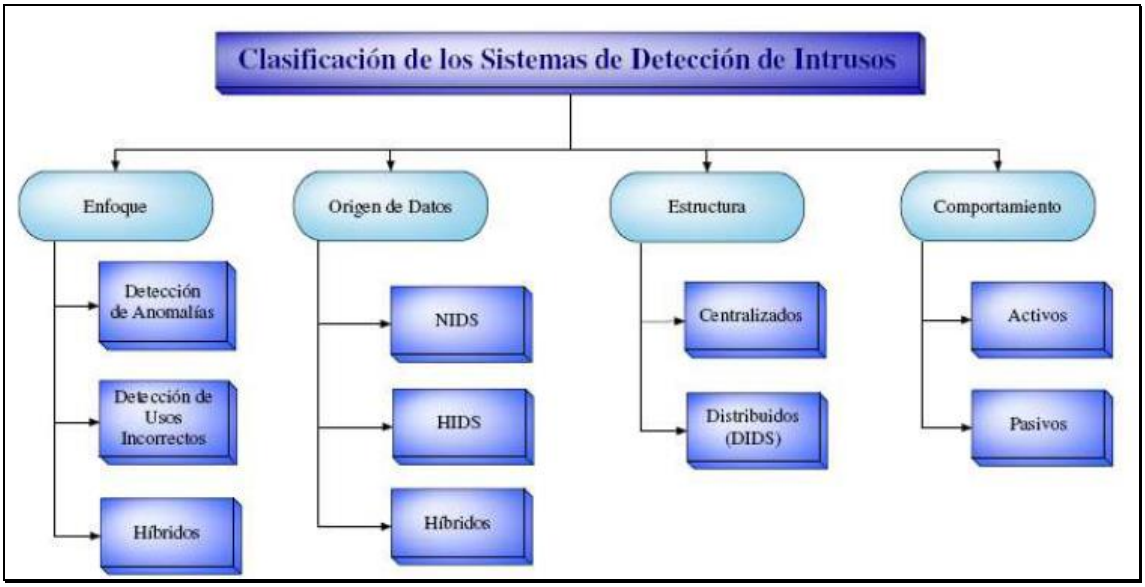


Figura 1- Clasificación de los Sistemas de Detección de Intrusos

He encontrado una web donde hace una interesante comparativa entre los principales IDS Opensource que existen y que deja muy bien parado a Bro IDS (Zeek): "*Bro es tanto un IDS basado en anomalías como en firmas. El tráfico capturado generará una serie de eventos. Por ejemplo, un evento podría ser un inicio de sesión de usuario a un FTP, conexión a servicio web o casi cualquier cosa…. además es capaz de detectar más patrones de actividad que la mayoría de IDS."*. Esta información la podemos encontrar en:

https://protegermipc.net/2017/02/22/mejores-ids-opensource-deteccion-de-intrusiones/

La verdad es que blogs, instalaciones y, en general, en número de resultados en productos como Snort y Suricata es bastante amplia, en el caso de Zeek es algo menor pero he podido encontrar información muy útil e interesante para la instalación e integración de Bro IDS (actualmente Zeek) con ELK Stack y que se puede consultar en los siguientes enlaces:

- https://logz.io/blog/bro-elk-part-1/
- https://logz.io/blog/bro-elk-part-2/

Incluso he podido encontrar algunos vídeos interesantes que podrán ayudar en la instalación e integración:

- https://www.youtube.com/watch?v=-2MUKmndTAU

- https://www.youtube.com/watch?v=HNVk31C2g4s

Una investigación interesante está basada en que muchos ataques informáticos recientes se han lanzado en múltiples etapas para evadir la detección de los sistemas de detección de intrusos (IDS) existentes. Algunas etapas del ataque pueden parecer inocentes si se verifican por separado. La correlación de eventos activos (AEC), que recopila y correlaciona eventos de red sospechosos dentro de un sistema de detección de intrusos en la red (NIDS) y qué implementaron AEC sobre Bro NIDS [BIN06]. Podría ser una línea interesante de investigación pero la orientación de este trabajo es otra.

También he podido encontrar una tesis interesante sobre la comparación entre dos de los IDS más conocidos, Zeek y Snort, para determinar sus funcionalidades en entornos de pocos recursos como una Rasperry Pi llamado *"A Comparison of Intrusion Detection Systems in Home Networks"* de Mikael Beremar y John Fryland.

Como bien dice, para un entorno del hogar, puede ser muy interesante la utilización de elementos con pocos recursos pero si, como es nuestro caso, nos movemos en entornos empresariales con gran volumen de tráfico necesitaremos equipos con algo más de capacidad aunque muchos de los conceptos ya investigados y puestos en funcionamiento pueden ser aprovechados a lo largo de este trabajo.

Para afrontar este proyecto, voy a utilizar recursos ya disponibles en el hospital o cuyo valor residual es prácticamente cero:

2 x Equipo Sobremesa → Valor residual 0 €

Destinado a la instalación de Debian 9 + Bro IDS → Coste licenciamiento 0 €

- HP COMPAQ 6000 PRO

- CORE 2 DUO E8400

- 4GB RAM DDR3

- 250GB Hd

- 1 x red - Ethernet 10Base-T/100Base-TX/1000Base-T - RJ-45 Intel 82567LM GbE Network Connection

1 x Servidor Virtual → Coste 0 €

Destinado a la instalación de Windows 2012 Server R2 (Licencia Datacenter) + ELK Stack. Al ser un servidor virtual podremos asignarle los recursos necesarios y de forma prácticamente dinámica sin necesidad se comprar equipamiento adicional.

3. Metodología y Planificación

3.1. Enfoque y método seguido

El enfoque y la metodología que se van a seguir durante la elaboración de este trabajo vendrán marcados por la consecución de los objetivos para cada una de las diferentes fases.

Para la definición de la planificación y objetivos del trabajo se centrará en la recopilación y lectura de información con lo que será un enfoque eminentemente teórico y dónde cobra especial importancia la consulta de documentación sobre la elaboración de trabajos de investigación y, más concretamente, de Grado y Fin de Máster.

En la fase de recopilación de información y decisión sobre las herramientas y elementos que utilizaremos para la implementación de la solución elegida, nos centraremos en la consulta de información para tomar una decisión de implementación con el mejor coste-beneficio posible sin dejar de lado el objetivo principal definido anteriormente.

Durante el desarrollo de la implementación y puesta en marcha nos centraremos en la definición del conjunto de medidas a adoptar para la obtención de resultados relevantes para la detección de actividades sospechosas. Para ello, es necesario identificar desde la experiencia, las fuentes externas y las herramientas implementadas qué actividades entran dentro del rango de "actividad normal", o se detectan directamente como actividad normal, y cuáles se salen de ese rango, o se detectan directamente como actividad anormal o sospechosa, principalmente utilizando ejemplos de detección.

En la fase de conclusiones, se intentará obtener una visión objetiva de los resultados obtenidos en las fases anteriores y que sea una base para posteriores trabajos en el mismo o similar ámbito.

3.2. Lista de Tareas

Para la elaboración del Trabajo de Fin de Máster se ha previsto una serie de tareas agrupadas por las distintas fases del mismo.

Plan de Trabajo

Esta primera fase está dividida en varias tareas que serán la base sobre la que podremos conseguir los objetivos finales.

Lectura del problema a resolver y tomas de contacto

Los primeros días se hará una lectura del enunciado, documentación relativa a cómo realizar un trabajo de fin de máster y a la lectura de algunos ejemplos que ayuden a la definición de las tareas siguientes.

Definición de objetivos

Una de las principales tareas a realizar será la definición de los objetivos que se pretende conseguir, de forma realista y teniendo en cuenta el tiempo y los recursos de los que dispondremos.

Definición de propuesta de metodología

Será necesario establecer el enfoque que le daremos a cada una de las fases del trabajo y la metodología que utilizaremos.

Planificación

Una correcta planificación es esencial en la consecución final de los objetivos, para ello es necesario identificar de forma correcta:

- Definición de tareas.- Lista de tareas a realizar.

- Cálculo tiempos de entrega.- Debemos establecer los tiempos, lo más aproximado posible, para cumplir con cada uno de los hitos establecidos por el planning de la asignatura.

- Identificación de riesgos.- Otro aspecto a destacar es la identificación de los riesgos que nos podemos encontrar a lo largo del desarrollo de cualquier trabajo ya que puede incrementar de manera sustancial el cumplimento de alguna de las tareas definidas o complicar la consecución de los objetivos marcados.

- Definición de costes y recursos.- Cualquier solución a un problema debe tener un coste que no exceda los beneficios que se pueda obtener por la implantación de la misma y para su justificación.

Finalización Entrega 1

La entrega del Plan de Trabajo será el primer hito.

Recopilación de Información, Análisis y Planteamiento

Durante esta fase se recopilará toda la información posible, se analizará y ayudará en la toma de decisión para el conjunto de herramientas que formarán parte de la solución completa a implementar.

Información y estudio herramientas IDS

Se realizará la recopilación de información de distintas herramientas de detección de intrusiones, profundizando en aquella que hayamos elegido finalmente. En nuestro caso será Zeek (Bro IDS).

Información y estudio Bases Datos NoSQL

Se realizará la recopilación de información de distintas herramientas de almacenamiento de información y que puedan tener una alta capacidad de integración con el IDS elegido en la fase anterior. Además, también tendremos en cuenta los complementos adicionales a la base de datos tanto para explotación como para representación y ayuda en toma de decisiones. En este caso será ELK Stack.

Machine Learning

La integración de todos los datos se realizará en ElasticSearch y para ayudar en la gestión de la seguridad y los eventos relacionados con la seguridad surge la necesidad de utilizar tecnología Machine Learnig que dispone el propio ELK Stack. En esta tarea, recopilaremos información para refrescar los conceptos de Machine Learnig y como pueden ayudarnos en la detección de intrusiones.

Redacción y Entrega 2

El siguiente hito nos marca la redacción de los pasos seguidos en las tareas anteriores y la entrega del estado del trabajo hasta ese momento.

Implementación

En esta fase se procederá a la implementación completa de la solución elegida y comprende las siguientes tareas:

Aprovisionamiento de recursos

Tanto de recursos hardware necesarios como descarga del software de sistemas operativos y de las herramientas necesarias a instalar.

Instalación, configuración e integración

En esta fase se procederá a la instalación y configuración de:

- Sistemas Operativos
- Zeek
- ELK Stack

Y por último se procederá a realizar las distintas integraciones IDS-ELK.

Test y puesta en marcha

Toda vez que todo el software está completamente instalado e integrado se procederá a:

- La definición y realización de batería de pruebas.
- La configuración definitiva reglas IDS, en caso que hayamos detectado faltantes.
- Puesta en productivo, recopilación de datos y monitorización.
- Análisis y detecciones.

Redacción y Entrega 3

Es necesario ir elaborando la documentación de la implantación conforme se va realizando y, por último, se procederá a la entrega de la misma, marcando otro hito más en el desarrollo de este trabajo de fin de máster.

Memoria Final

Comprende las siguientes tareas:

Presentación y defensa del TFM

La memoria final es el elemento primordial que debe ser entregado como justificación de todo el trabajo realizado.

- Conclusiones.- Una vez realizada la entrega 3 y obtenidos y analizados los datos procederemos a la redacción de las conclusiones.

- Elaboración de Memoria Final.- Una revisión completa de toda la documentación elaborada y una remaquetación final seguramente será necesaria.

Entrega 4

La entrega de la Memoria Final marca tanto un hito como uno de los principales objetivos de todo TFM.

Elaboración y presentación vídeo TFM

En la UOC es necesaria la elaboración de un video explicativo de unos 15 minutos de duración para el que tendremos que aprovisionarnos del software y hardware necesario para la realización del mismo.

Defensa del TFM

La defensa del trabajo de fin de máster marca el final de la asignatura y que consistirá en la respuesta a las preguntas realizadas por el tribunal que evalúa el trabajo realizado.

3.3. Identificación de Riesgos

A continuación, se detallarán los principales riesgos, en distintos ámbitos, que pueden afectar en mayor o menor medida al correcto desarrollo de este trabajo. Será necesario tanto la identificación como aquellas acciones que podremos realizar para mitigar, en la medida de lo posible, los efectos en la planificación.

Riesgo 1.- Sobredimensión del ámbito del trabajo

La definición inicial del proyecto o de los objetivos a alcanzar son más amplios que la dedicación o calendario previstos.

Probabilidad / Impacto (1-5): 3 / 5

Acciones mitigadoras:

- Descartar algunos temas que se desvíen del objetivo principal del trabajo, eliminar tareas superfluas o secundaria y, en último caso, reducir el ámbito o algunos objetivos secundarios.

Riesgo 2.- Falta de documentación de apoyo

Que la documentación de apoyo inicial para la implementación/integración de la solución completa no sea suficiente.

Probabilidad / Impacto (1-5): 2 / 4

Acciones mitigadoras:

- Ampliar documentación en internet, bibliotecas virtuales, otros trabajos con líneas de investigación similar, vídeos y foros de seguridad.

Riesgo 3.- Obtención de resultados escasos o poco relevantes

La batería de pruebas/tests, configuración de reglas u obtención de tráfico de red insuficiente.

Probabilidad / Impacto (1-5): 2 / 5

Acciones mitigadoras:

- Redacción clara de los principales riesgos a los que nos podremos enfrentar y documentarse y redactar los resultados obtenidos para las distintas posibilidades de conexión de nuestro sniffer de red.

- Tener preparada entorno de test con máquina de hacking ético (tipo Kali Linux) para su utilización y comprobación de batería de pruebas de test.

Riesgo 4.- Falta de tiempo para redactar la memoria

Si nos centramos en la consecución de cada uno de los hitos del proyecto, puede que no se llegue a tiempo para la redacción de la memoria o que la elaboración de la misma implique más tiempo del planificado.

Probabilidad / Impacto (1-5): 2 / 5

Acciones mitigadoras:

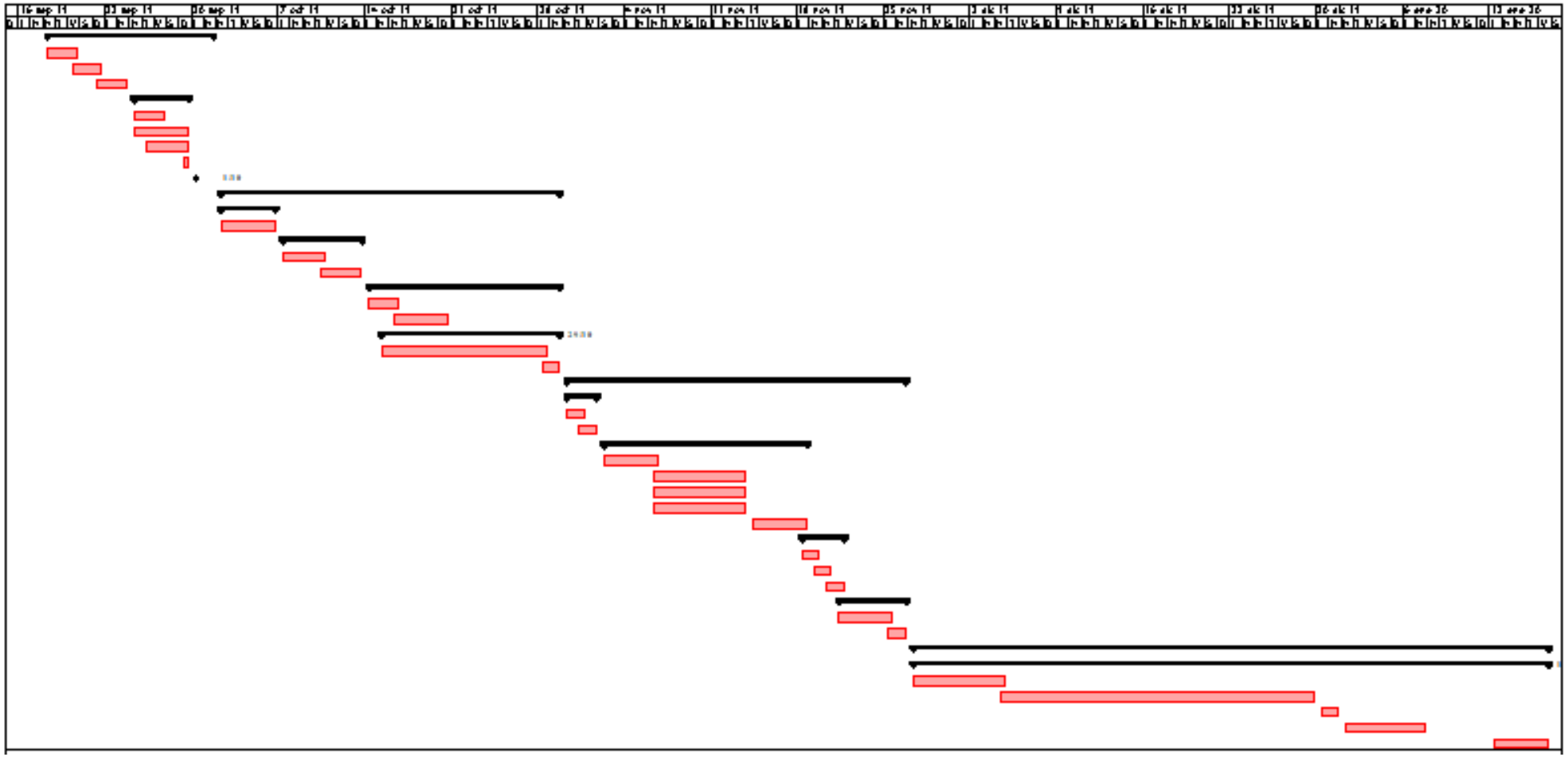
- Ir documentando cada fase del proyecto conforme se van realizando avances más o menos significativos dejando únicamente para el final las conclusiones obtenidas y la revisión de formato.

3.4. Planificación del Trabajo

Lista de tareas con planificación:

		Nombre	Trabajo	Inicio	Terminado	Marca tareas como hito
1		**PLAN DE TRABAJO**	**26 horas**	**18/09/19 8:0**	**1/10/19 17:00**	☐
2		Lectura del problema a resolver y toma	3 horas	18/09/19 8:00	20/09/19 17:00	☐
3		Definición de objetivos	3 horas	20/09/19 8:00	22/09/19 17:00	☐
4		Definición de propuesta de metodologí	3 horas	22/09/19 8:00	24/09/19 17:00	☐
5		**Planificación**	**15 horas**	**25/09/19 8:0**	**29/09/19 17:.**	☐
6		Definición de tareas	6 horas	25/09/19 8:00	27/09/19 17:00	☐
7		Cálculo tiempos de entrega	4 horas	25/09/19 8:00	29/09/19 17:00	☐
8		Identificación de riesgos	3 horas	26/09/19 8:00	29/09/19 17:00	☐
9		Definición de costes y recursos	2 horas	29/09/19 8:00	29/09/19 17:00	☐
10		Finalización Entrega 1	2 horas	30/09/19 8:00	1/10/19 17:00	☑
11		**RECOPILACIÓN DE INFORMACIÓN,**	**54 horas**	**2/10/19 8:00**	**29/10/19 17:.**	☐
12		**Información y estudio herramient**	**10 horas**	**2/10/19 8:00**	**6/10/19 17:00**	☐
13		Estudio Bro IDS	10 horas	2/10/19 8:00	6/10/19 17:00	☐
14		**Información y estudio Bases Dato**	**14 horas**	**7/10/19 8:00**	**13/10/19 17:.**	☐
15		Estudio Elastic Search	8 horas	7/10/19 8:00	10/10/19 17:00	☐
16		Estudio Kibana+Addons integración	6 horas	10/10/19 8:00	13/10/19 17:00	☐
17		**Información y estudio SIEMs**	**30 horas**	**14/10/19 8:0**	**29/10/19 17:.**	☐
18		Revisión y estudio conceptos Machine	4 horas	14/10/19 8:00	16/10/19 17:00	☐
19		Análisis, comparativa y elección de SI	8 horas	16/10/19 8:00	20/10/19 17:00	☐
20		**Redacción y Entrega 2**	**18 horas**	**15/10/19 8:0**	**29/10/19 17:.**	☑
21		Elaboración documentación Entrega	15 horas	15/10/19 8:00	28/10/19 17:00	☐
22		Revisión y entrega Entrega 2	3 horas	28/10/19 8:00	29/10/19 17:00	☐
23		**IMPLEMENTACIÓN**	**51 horas**	**30/10/19 8:0**	**26/11/19 17:.**	☐
24		**Aprovisionamiento de recursos**	**6 horas**	**30/10/19 8:0**	**1/11/19 17:00**	☐
25		Recursos hardware	2 horas	30/10/19 8:00	31/10/19 17:00	☐
26		Recursos software	4 horas	31/10/19 8:00	1/11/19 17:00	☐
27		**Instalación, configuración e integ**	**27 horas**	**2/11/19 8:00**	**18/11/19 17:.**	☐
28		Instalación y configuración de Sistema	6 horas	2/11/19 8:00	6/11/19 17:00	☐
29		Instalación y configuración IDS BRO	5 horas	6/11/19 8:00	13/11/19 17:00	☐
30		Instalación y configuración ELK Stack	6 horas	6/11/19 8:00	13/11/19 17:00	☐
31		Instalación y configuración SIEM	5 horas	6/11/19 8:00	13/11/19 17:00	☐
32		Integraciones IDS-ELK-SIEM	5 horas	14/11/19 8:00	18/11/19 17:00	☐
33		**Test y puesta en marcha**	**8 horas**	**18/11/19 8:0**	**21/11/19 17:.**	☐
34		Definición y realización de batería de	3 horas	18/11/19 8:00	19/11/19 17:00	☐
35		Configuración definitiva reglas IDS	2 horas	19/11/19 8:00	20/11/19 17:00	☐
36		Puesta en marcha y monitorización	3 horas	20/11/19 8:00	21/11/19 17:00	☐
37		**Redacción y Entrega 3**	**10 horas**	**21/11/19 8:0**	**26/11/19 17:.**	☑
38		Documentación de la implantación	8 horas	21/11/19 8:00	25/11/19 17:00	☐
39		Entrega 3	2 horas	25/11/19 8:00	26/11/19 17:00	☐
40		**MEMORIA FINAL**	**43 horas**	**27/11/19 8:0**	**17/01/20 17:.**	☐
41		**Presentación y defensa del TFM**	**43 horas**	**27/11/19 8:0**	**17/01/20 17:.**	☑
42		Conclusiones	10 horas	27/11/19 8:00	4/12/19 17:00	☐
43		Elaboración de Memoria Final	20 horas	4/12/19 8:00	29/12/19 17:00	☐
44		Entrega 4	2 horas	30/12/19 8:00	31/12/19 17:00	☐
45		Elaboración y presentación video TFM	6 horas	1/01/20 8:00	7/01/20 17:00	☐

Diagrama de Gantt:



4. Recopilación de información, análisis y planteamiento

4.1. Herramientas IDS

Definición de IDS

Un IDS (Sistema de Detección de Intrusiones) es un elemento que escucha y analiza toda la información que circula por una red de datos e identifica posibles ataques. Los IDS están siendo usados ampliamente a lo largo de estos últimos años por muchas compañías ya que proporcionan una capa adicional de seguridad y ayudan a entender el ataque, estimar los daños causados y tratan de prevenir otros ataques similares.

Las primeras investigaciones sobre detección de intrusos comienzan en 1980 en un trabajo de consultoría realizado para el gobierno norteamericano por James P. Anderson [AND72]. Anderson, expuso la idea de que el comportamiento normal de un usuario podría caracterizarse mediante el análisis de su actividad en los registros de auditoría. Así, los intentos de intrusión podrían descubrirse detectando actividades anómalas que se desviaran notablemente de su comportamiento normal. A partir de este momento, comienzan una gran variedad de investigaciones en busca de técnicas y algoritmos para el análisis del comportamiento de un sistema informático.

Los sistemas de detección de intrusos suelen estar formados por:

- ✓ Sensores: Tienen la responsabilidad de coleccionar datos de interés.
- ✓ Analizadores: Reciben la información recogida por los sensores y determinan si está ocurriendo o ha ocurrido una intrusión.

En el apartado del arte, se introdujo una figura donde se indican la clasificación de los distintos IDS que de forma general se pueden clasificar:

- **Según su enfoque**
  - **Basados en detección de anomalías**, técnicas estadísticas y/o de aprendizaje automático para distinguir el comportamiento usual del anormal
  - **Basados en usos incorrectos** que trabajan en la detección de uso indebido comparando firmas con la información recogida.
  - **Sistemas híbridos**: Tanto los IDS basados en detección de usos incorrectos como los basados en detección de anomalías, presentan ventajas e inconvenientes que hacen que ninguna de las dos soluciones sea claramente superior a la otra. Así, los IDS basados en firmas resultan más fiables y proporcionan mejores rendimientos frente a ataques conocidos. Sin embargo, no presentan capacidad para detectar nuevos ataques que no se encuentren en la base de datos de firmas. Por el

contrario, los IDS basados en anomalías presentan la capacidad de detectar ataques previamente desconocidos, aunque su rendimiento resulte inferior y con otros problemas de falsos positivos.

- **Según el origen de los datos**
    - **Sistemas de detección de Intrusos basados en Host (HIDS):** Están diseñados para monitorizar, detectar y responder a los datos generados por un usuario o un sistema en un determinado host, fueron los primeros IDS que surgieron.
    - **Sistemas de detección de Intrusos de Red (NIDS):** Los NIDS son muy parecidos a los HIDS en tanto que monitorizan, detectan y responden ante los datos generados, pero en vez de proteger un determinado host, reciben los datos de la red local donde estén instalados. Tienen la ventaja de ser, normalmente, pasivos, de forma que no interfieren en la carga de la red. Actúan mediante la utilización de un dispositivo de red configurado en modo promiscuo, capturando todos los paquetes que pasan por él y almacenando la información de estos paquetes en un repositorio para su posterior análisis. No detectan los ataques locales (con acceso físico) de los usuarios a un determinado host.
    - **IDS Híbridos:** Se instalan sensores en cada host que permiten la detección local y un sensor en cada segmento de red a vigilar. Así cubrimos las necesidades completas de detección y explotamos las ventajas de ambas arquitecturas.
- **Según su estructura**
    - **Distribuidos:** Sensores distribuidos por diversas máquinas y puntos de red y sólo se centraliza la información relevante.
    - **Centralizados:** Sensores distribuidos pero que envían toda la información a un único nodo central.
- **Según su comportamiento**
    - **Pasivos (IDS):** Son aquellos que sólo notifican o guardan la información relevante pero no actúan sobre el ataque o atacante.
    - **Activos (IPS):** Son aquellos que sí actúan con diversas acciones que pueden interferir en el ataque como bloqueando la IP, activando determinadas reglas del firewall, etc.

Hay que destacar que para cada una de las técnicas de detección de anomalías y de uso indebido se utilizan diferentes tipos de algoritmos para dotar al sistema de conocimiento y de aprendizaje autónomo.

La detección de anomalías se basa en la suposición de que el comportamiento que tienen los usuarios y la red es lo suficientemente regular como para obtener un patrón de comportamiento, de forma que cualquier desviación significante pueda ser considerada como una evidencia de una intrusión en el sistema. Posibles inconvenientes:

- Este tipo de sistemas son muy propensos a los falsos positivos; esto es, se dispara una alerta aunque la actividad que está marcando sea normal y permitida en nuestra red.

- Se tarda mucho en que un IDS de detección de anomalías desarrolle un modelo preciso de la red.

- Al principio, el sistema generaría tantas alertas que sería casi inútil.

- Depende de la calidad del proceso de aprendizaje: un entrenamiento demasiado restringido podría causar un sistema con un alto porcentaje de falsos positivos, mientras que un entrenamiento demasiado general, puede producir un alto porcentaje de falsos negativos.

IDS en la actualidad

Para la elección de un determinado IDS, he indagado en distintos sitios entre los que destaca http://www.comparitech.com. Una empresa fundada en 2015, con un equipo de 30 investigadores, escritores, desarrolladores y editores que cubren una amplia gama de servicios en línea, incluidos VPN, administradores de contraseñas, protección contra robo de identidad, antivirus, proveedores de Internet, monitoreo de redes y más. Con más de 1 millón de visitas mensuales.

Donde se aborda un artículo llamado *"2019 Best Intrusion Detection Systems"* donde Stephen Cooper nos hace la siguiente tabla comparativa:

IDS	HIDS/NIDS	Unix	Linux	Windows	Mac OS	Free
SolarWinds Security Event Manager	Both	No	No	Yes	No	No
Snort	NIDS	Yes	Yes	Yes	No	Yes
OSSEC	HIDS	Yes	Yes	Yes	Yes	Yes
Suricata	NIDS	Yes	Yes	Yes	Yes	Yes
Bro	NIDS	Yes	Yes	No	Yes	Yes
Sagan	Both	Yes	Yes	No	Yes	Yes
Security Onion	Both	No	Yes	No	No	Yes
AIDE	HIDS	Yes	Yes	No	Yes	Yes
Open WIPS-NG	NIDS	No	Yes	No	No	Yes
Samhain	HIDS	Yes	Yes	No	Yes	Yes
Fail2Ban	HIDS	Yes	Yes	No	Yes	Yes

Figura 2- Comparativa IDS 2019

La elección del IDS intervendrán varios factores:

- Coste del licenciamiento

- Adecuación a las necesidades de detección del trabajo

- Documentación disponible y popularidad

De esta forma, procedemos de la siguiente manera:

1.- Descartamos la primera opción ya que no es gratuita y sólo podemos disponer de una versión trial de 30 días.

2.- Descartamos todos los que son sólo Host-Based Detection.

3.- Descartamos aquellos cuyo volumen de entradas en Google son menores.

Así, sólo nos quedaría SNORT y Bro IDS (Zeek). Ambos cubrirían los requerimientos anteriores y tienen más de 3 Millones de entradas en Google.

Por último, me he decantado por Zeek ya que si bien el objetivo de éste es el mismo que el de Snort, el modelo aplicado para la resolución de las amenazas en un entorno de red es completamente distinto.

4.2. Zeek

Bro, que pasó a llamarse Zeek a finales de 2018, está basado en firmas y en anomalías. Su motor de análisis convertirá el tráfico capturado en una serie de eventos. Un evento podría ser un inicio de sesión de usuario en FTP, una conexión a un sitio web o prácticamente cualquier cosa. El poder del sistema es lo que viene después del motor de eventos y ese es el intérprete de scripts de política. Este motor de políticas tiene su propio lenguaje (Bro-Script) y puede realizar algunas tareas muy potentes y versátiles.

Como se mencionaba antes, los IDS basados en reglas utilizan un conjunto de patrones que sirven para detectar amenazas o incluso ataques que se están llevando a cabo en la red. Dichos patrones pueden ser muy flexibles y tener unos muy buenos niveles de efectividad contra amenazas comunes, pero dado el dinamismo del tráfico en segmentos de red y la facilidad que tienen los atacantes de ocultar payloads, malware y todo de tipo de paquetes maliciosos, es muy probable que un sistema basado en reglas sea capaz de filtrar todas las posibles amenazas que pueden producirse.

En este sentido, los Zeek supone un enfoque distinto a los IDS tradicionales basados en reglas, ya que además de capturar eventos potencialmente peligrosos, provee de un framework bastante potente para analizar todos los eventos y el tráfico generado, de tal forma que el analista de seguridad podrá verificar si efectivamente si se está llevando a cabo un ataque o si hay algún tipo de amenaza en la red.

Las principales características de Zeek, recogidas de su página oficial, son:

- Adaptable: Al disponer de un lenguaje de script propio, permite política de monitoreo específicas.

- Eficiente: Apunta a redes de alto rendimiento y se usa operacionalmente en una variedad de grandes sitios.

- Flexible: No está restringido a ningún enfoque de detección en particular y no se basa en firmas tradicionales.

- Forense: Registra exhaustivamente lo que ve y proporciona un archivo de alto nivel de la actividad de una red.

- Análisis en Profundidad: Por defecto viene con analizadores para muchos protocolos, lo que permite el análisis semántico de alto nivel en la capa de aplicación.

- Con estado a Alto Nivel: Mantiene un amplio estado en la capa de aplicación sobre la red que monitoriza.

- Interfaces Abiertas: Interactúa con otras aplicaciones para el intercambio de información en tiempo real.

- Open Source: Viene con Licenciamiento BSD, permitiendo uso gratuito sin restricciones virtuales.

Otras características destacables de Zeek

Como analista, Zeek puede automatizar parte del trabajo, puede descargar archivos vistos en la red y enviarlos para análisis de malware, notificar si se encuentra un problema, poner en una lista negra la fuente y apagar el equipo del usuario que lo descargó, Puede realizar un seguimiento de los patrones de uso de un usuario después de haber contactado una IP de una base de datos de reputación, etc.

Si no se es analista, esta herramienta tendrá una curva de aprendizaje desafiante. Dado que se desarrolló como una herramienta de investigación, inicialmente no se centró en cosas como GUI, usabilidad y facilidad de instalación. Si bien hace muchas cosas interesantes fuera de la caja, muchas de esas cosas no son procesables de inmediato y pueden ser difíciles de interpretar.

No dispone de una GUI nativa, pero hay herramientas de código abierto de terceros disponibles para un front-end web para consultar y analizar alertas provenientes de Bro-IDS como ELK Stack.

Es de destacar la cantidad de tipos de logs generados por Zeek y que incluiremos en el ANEXO II del trabajo y que se ha obtenido de la página oficial de Zeek (http://www.zeek.org). Junto con estos logs, es posible generar, mediante Scripting de la propia herramienta, nuevos logs para futuras necesidades aunque esta característica, muy interesante si no se cumplen todas las necesidades de monitorización, se escapa del ámbito de este trabajo.

4.3. Bases Datos NoSQL

Los sistemas de detección de intrusos tienen estas particularidades:

- Generan registros muy valiosos con detalles de uso de red y de alertas.

- Recopilan gran cantidad de datos que necesitan almacenes con estructuras con una gran cantidad de campos.

- En la mayoría de ocasiones, no se sabe de antemano qué eventos habrá que almacenar ni qué estructura será necesaria.

- Se necesita dar sentido a tantos datos y hacer que dicha información sea procesable para análisis avanzados, alertas y funciones de búsqueda.

Existen herramientas de sistema de gestión de logs o registros pero traspasar y analizar la información obtenida por un IDS puede ser un gran desafío. Por ello, desde hace años son muchos los que apuestan por utilizar bases de datos NoSQL [WYIL12] [DUS17]:

- No tienen esquemas.

- No usan SQL como el principal lenguaje de consultas.

- No garantizan la propiedad ACID (atomicidad, coherencia, aislamiento y durabilidad).

- Los datos almacenados no requieren estructuras fijas como tabla

- Escalan bien horizontalmente aunque hacen uso amplio de la memoria principal del computador.

- Resuelven el problema de los altos volúmenes de información y la inmensa cantidad de consultas y transacciones diarias.

En resumen, no son relacionales.

En nuestro caso, se va a capturar gran cantidad de información procedente, para este proyecto, de una única fuente (Bro IDS) aunque, aprovechando la escalabilidad que nos ofrecen las Bases de Datos NoSQL, se irá ampliando el número de sondas y fuentes para aumentar la probabilidad detección de anomalías o intrusiones que puedan producirse en nuestra red.

Entre las más populares bases de datos NoSQL tenemos MongoDB y ElasticSearch. A continuación, realizaremos una comparativa entre ambas teniendo en cuenta tanto sus puntos fuertes como sus debilidades [COI19].

Adaptabilidad JSON

MongoDB es considerado como uno de los enfoques tradicionales para una base de datos de documentos. Es capaz de manejar documentos JSON en sus colecciones. También es posible convertirlos a todos en BSON. Un BSON no es más que una versión binaria de JSON. Su objetivo principal es ayudar a los usuarios a reducir los problemas que llegan durante los campos de tamaño arbitrario.

ElasticSearch, si se compara con MongoDB, tiene una excelente biblioteca de búsqueda que facilita a los usuarios administrar su tarea con mucha facilidad. Al igual que MongoDB, también es capaz de manejar documentos JSON en los índices. Sin embargo, debe tenerse en cuenta que, en el caso de ElasticSearch, no es posible la conversión binaria. Por lo tanto, MongoDB tiene una ventaja aquí, pero en el caso de Elasticsearch, es posible analizar el contenido o los datos presentes en un documento. Los índices se crean después del análisis, lo que elimina muchos errores. Básicamente, los valores de varios campos se toman para una comprensión efectiva.

Comparación en términos de excavación de datos

Si ambos se comparan en términos de excavación de datos, ElasticSearch es una buena opción. En realidad, es capaz de excavar datos en diferentes métodos y así garantizar la fiabilidad. Como es conocido, el número total de Índices es extremadamente importante en cualquier base de datos. Los índices se actualizan cada vez que se inserta un registro nuevo. La complejidad de las tareas depende completamente del tamaño de los índices. ElasticSearch enfrenta este problema en gran medida mientras que MongoDB no. Esto se debe a que generalmente contiene más índices.

Ésto te hace pensar que ElasticSearch es una buena opción para considerar como tu base de datos pero se debe tener en cuenta que la mayoría de las aplicaciones en el escenario actual no necesitan índices en gran medida. MongoDB tiene índices limitados en comparación con ElasticSearch, por ello es bastante rápido y confiable. Por lo tanto, puede continuar con ellos en caso de que el tamaño de la aplicación no sea demasiado grande.

Problema de pérdida de datos en ElasticSearch

El mayor problema con Elasticsearch es que muchos expertos han informado que pierde operaciones de escritura. Esto sucede a menudo mientras se reforman los grupos y por ello ElasticSearch no se considera una buena opción para la base de datos principal.

De nuevo, las consideraciones anteriores te confunden al decidir qué base de datos considerar así que resumiendo un poco: Los índices MongoDB son buenos, pero su efecto en el rendimiento general del Índice no se considera el mejor. Aquí, ElasticSearch es considerada como una de las mejores opciones. Es porque se puede confiar para mantener los índices simples y mínimos. Todos los registros se pueden guardar fácilmente y los cambios son posibles en cualquier momento. En ambas tecnologías, no es necesario que defina la estructura de una base de datos a la hora de realizar cualquier forma de modificación de la información.

Búsqueda de datos

A veces, los datos son tan voluminosos que las organizaciones se enfrentan a muchos problemas cuando tienen que encontrar un documento o registro específico. MongoDB está equipado con una búsqueda geoespacial que lo hace simplemente el mejor para encontrar los datos que se requieren exactamente y esto se debe a que es capaz de almacenar datos en formato BSON.

Sin embargo, no tiene ninguna interfaz para API REST, que es su inconveniente en comparación con ElasticSearch. Hay varias funcionalidades para encontrar los documentos relacionados que no se pueden realizar con MongoDB. En cambio, en varios casos, solo puede especificar el idioma.

No hay posibilidad de transacciones ACID con MongoDB. Los documentos no pueden relacionarse mediante un método paralelo. Los desarrolladores siempre tienen un desafío clave para encontrar una relación entre los documentos y, por ejemplo, realizar las consultas múltiples. Esto a veces consume mucho tiempo y especialmente cuando se emplean aplicaciones a gran escala.

ElasticSearch también es capaz de realizar una búsqueda avanzada, pero a veces carece de escalabilidad. Su característica única es la disponibilidad de datos, incluso si los nodos están apagados por algún motivo. En realidad, almacena los datos automáticamente y los usuarios no tienen que preocuparse por esto. ElasticSearch es capaz de manejar consultas a través de la API REST y ésta es su ventaja sobre MongoDB. Los documentos planos se pueden almacenar fácilmente y sin degradar el rendimiento de toda la base de datos. Además de esto, ElasticSearch es capaz de manejar datos a través de filtros.

Ya disponemos de suficiente información sobre MongoDB y ElasticSearch y podremos elegir la base de datos primaria en su organización. En la mayoría de los aspectos, ambos enfoques son poderosos y fáciles de considerar.

Decisión

En definitiva, y resumiendo:

- Son tecnologías diferentes que, en general, se usan para casos de uso distintos.

- MongoDB es una base de datos NoSql documental (almacena fragmentos de datos en formato JSON) y ElasticSearch es un motor de indexación de datos NoSql (basado en el motor Lucene).

- MongoDB es entonces un motor de base de datos de propósito general mientras que Elastic es un motor de indexación de textos planos distribuido. Por ejemplo, Elastic posee un modelo transaccional de consistencia eventual y no soporta el concepto de transacción en sus APIs ni tiene concepto de usuarios y permisos, lo que no ocurre con MongoDB.

- El escenario de uso típico de Elastic es “ingerir” grandes cantidades de datos de diferentes fuentes para hacer análisis de datos y correlaciones, calcular indicadores y presentar información de forma sumarizada.

El principal motivo por el que nos decantamos por Elastic Search es el último que indicamos, la capacidad que tiene de “ingerir” datos de multitud de fuentes distintas ya que, aunque en un principio el trabajo está orientado a la obtención de una fuente concreta, nos permitirá ampliar la recolección de información de otros productos, fuentes, sondas, etc. en un futuro para la detección de otras posibles amenazas o anomalías en un futuro con la utilización del paquete Elastic Stack [ABD15] [SUN17].

4.4. SIEMs

Gartner introdujo el término “SIEM" en un informe del año 2005 titulado “Mejore la Seguridad de IT con la Gestión de Vulnerabilidades” [AMR05].

La tecnología SIEM surgió de la combinación de las funciones de dos categorías de productos: SEM (gestión de eventos de seguridad) y SIM (gestión de información de seguridad).

- ✓ SEM centraliza el almacenamiento y permite un análisis casi en tiempo real de lo que está sucediendo en la gestión de la seguridad, detectando patrones anormales de accesibilidad y dando mayor visibilidad a los sistemas de seguridad.

- ✓ SIM recopila los datos a largo plazo en un repositorio central para luego analizarlo, proporcionando informes automatizados al personal de seguridad informática.

Ambas funciones permiten que se pueda actuar más rápidamente sobre los ataques, ya que por un lado ofrecen más visibilidad y por otro permiten utilizar los datos para la supervisión y el análisis de la seguridad en tiempo real, avisando de los ataques que se están produciendo, o incluso los que se van a producir.

Muchos de los actuales sistemas SIEM trabajan o se apoyan en tecnología de Machine Learnig para poder gestionar de forma proactiva las potenciales vulnerabilidades, protegiendo de devastadoras filtraciones de datos.

Un SIEM es una herramienta fundamental para la seguridad donde se pueden centralizar todos los incidentes de seguridad con millones de entradas provenientes de distintos sistemas y/o fuentes (antivirus, firewalls, NIDS, HIDS, ...) donde todo puede ser colocado en un formato común, listo para que la solución SIEM pueda ejecutar su análisis y correlación.

Esto reduce la carga de trabajo de los equipos de seguridad IT y le permite aprovechar una vista optimizada de las actividades y las potenciales preocupaciones.

4.5. Plataforma ELK

La Plataforma ELK es un conjunto de herramientas que al combinarse crean una plataforma robusta de administración de datos permitiendo el monitoreo, consolidación y análisis de los mismos.

Existen 3 herramientas principales que componen este sistema son:

- ElasticSearch: Es un motor de análisis y búsqueda distribuido basado en JSON.

- Logstash o Bean: Procesan y envían los datos ya procesados a Elasticsearch para ser indexados.

- Kibana: Interfaz gráfica de usuario de la plataforma ELK.

Y por las iniciales de estos (Elastic, Logstash y Kibana) el conjunto también es conocido como “stack ELK” o simplemente “ELK”.

Esta plataforma está disponible en 4 modalidades:

- Apache2: Software de código abierto, bajo la licencia Apache 2, salvo algunas funcionalidades adicionales que son distribuidas con licencia propietaria. Ofrece la oportunidad de ver, utilizar y hasta modificar las herramientas sin costo alguno, pero toda administración de las herramientas debe ser llevado a cabo personalmente.

- Basic: Ofrece multitud de funcionalidad con licencia gratis vitalicia (descritas en Anexo I) y seleccionada para la realización de este trabajo.

- Propietaria: La mayoría de las funcionalidades extra incluídas en el X-Pack, aunque algunas básicas ya se incluyan en la licencia Basic.

- Elastic Cloud: Como servicio de pago, Elastic Cloud pone a disposición todas las herramientas y las funcionalidades adicionales en servidores administrados por ellos.

Ventajas de la plataforma ELK

Hay muchas soluciones distintas para cubrir la necesidad de procesamiento, monitoreo y visualización de datos, tanto de pago como libres, pero lo que distingue a ELK sobre otras es principalmente:

- Potencia: Ofrece mucha funcionalidad con un bajo costo técnico, las configuraciones son mínimas para empezar.

- Escalabilidad: Elasticsearch es una herramienta diseñada para manejar terabytes de datos sin ningún problema. Su arquitectura le permite expandirse de forma rápida y fácil.

- Flexibilidad: La configuración es flexible y puede adaptarse a cualquier necesidad y entorno. Puede utilizarse para análisis de cualquier recogida de datos a gran escala.

- Extensible: Dispone de un ecosistema de extensiones (plugins) alrededor de sus herramientas que han creado un número importante de funcionalidades extras y gratuitas para facilitar el trabajo con ellas.

- Código Abierto: Con lo que ofrece ventajas competitivas sobre otras plataformas porque permite la rápida corrección de errores gracias a la comunidad, la creación de extensiones e incluso incremente la base de usuarios al permitir utilizar las herramientas sin requerir pago alguno, lo que incrementa el conocimiento compartido de las herramientas.

BEATS

Beats es la plataforma para los agentes de datos con un solo propósito. Envían datos de cientos o miles de máquinas y sistemas a Logstash o Elasticsearch.

Los Beats son excelentes para recopilar datos. Se quedan en tus servidores, con tus contenedores, o se despliegan como funciones, y luego centralizan los datos en Elasticsearch. Y si deseas una mayor potencia de procesamiento, Beats también puede enviar datos a Logstash para tareas de transformación y análisis.

Existen agentes para todo tipos de datos:

- Filebeat: Permite, mediante la interfaz de registros, seguir los archivos directamente en Kibana. Usa la barra de búsqueda para filtrar por servicio,

aplicación, host, centro de datos u otros criterios para rastrear el comportamiento curioso en sus registros agregados.

- Metricbeat: Recopila métricas de sistemas y servicios, desde la CPU a la memoria, de Redis a NGINX, y mucho más, Metricbeat es una forma liviana de enviar estadísticas de sistema y servicio.

- Packetbeat: Es un analizador de paquetes de red ligero que envía datos desde sus hosts y contenedores a Logstash o Elasticsearch.

- Winlogbeat: Trasmite los registros de eventos de Windows a Elasticsearch y Logstash.

- Auditbeat: Recopila los datos del marco de auditoría de Linux y monitorea la integridad de sus archivos. Auditbeat envía estos eventos en tiempo real al resto de Elastic Stack para su posterior análisis.

- Heartbeat: Monitorea los servicios para su disponibilidad con sondeo activo. Dada una lista de URL, Heartbeat hace la simple pregunta: ¿estás vivo? Heartbeat envía esta información y el tiempo de respuesta al resto de Elastic Stack para su posterior análisis.

- Functionbeat: Implementa como función en la plataforma de "función como servicio" (FaaS) de su proveedor de la nube para recopilar, enviar y monitorear datos de sus servicios en la nube.

En el ámbito de este trabajo nos centraremos en las funciones de Filebeat para la transformación y envío de los logs a Elasticsearch.

Kibana

Como se ha indicado anteriormente, Kibana es una herramienta de análisis y visualización de datos que nos será de vital importancia para ayudarnos en la detección de comportamientos anómalos con la información recopilada desde Zeek (Bro IDS) y cargada en nuestra base de datos NoSQL.

El tablero de Kibana ofrece varios diagramas interactivos, datos geoespaciales y gráficos para visualizar necesidades complejas. Se puede usar para buscar, ver e interactuar con los datos almacenados en los directorios de Elasticsearch. Kibana ayuda a realizar análisis de datos avanzados y visualizar sus datos en una variedad de tablas, gráficos y mapas.

Al acceder a Kibana podemos realizar varias acciones:

Descubrir (Discover)

Sirve para hacer una primera exploración de los datos. La pantalla de búsqueda está dividida en tres partes principales:

- Un buscador
- Un mapeo de campos
- Un espacio de resultados

Las principales acciones que se pueden realizar en cualquier elemento que se puede definir en Kibana, tanto para buscar como para configurar la visualización de la pantalla, son:

- Selector de índices: En el desplegable se encuentran los distintos índices importados en la instancia de Kibana, este desplegable nos permite mover a través de ellos. Incluso algunos filtros se mantienen entre índices si hay campos coincidentes.

- Buscador: Nos permite hacer toda una serie de preguntas sobre nuestra base de datos, una forma de ver si nuestra query está funcionado es comprobar el recuento de “hits” que aparece justo encima del buscador.

Otros aspectos importantes que tendremos que tener en cuenta para definir cada uno de los elementos que vamos a necesitar en Kibana son:

- Filtros: Los filtros gráficos pueden hacer más o menos las mismas operaciones de filtro que acabamos de ver en el buscador, con la ventaja de que se pueden sumar varios filtros con facilidad.

- Available fields: La barra lateral sirve para poder inspeccionar las cabeceras de los datos, dar una primera visión de los datos que contienen y configurar el panel de resultados.

- Configuración: La pequeña rueda que está al lado de “Availaible Fields” despliega una serie de opciones para que se muestren más o menos campos. En caso de datos no tabulares, como los de OCDS, se aconseja desmarcar “Hide missing fields”, para que se muestren los campos que están dentro de otros campos.

- Campos: Todos los campos van acompañados de un símbolo que identifica el tipo de datos que contiene, el reloj cuando es temporal, el numeral o tecla gato para identificar número, la «t» para identificar los textos o strings, una esfera mitad negra significa que el campo es un booleano, y el símbolo interrogante que desconoce qué tipo de campo tiene, normalmente será porque contiene a su vez más campos a su interior. Al clicar sobre un campo se desplegará un gráfico que mapea los primeros 500 valores del campo.

Kibana no sólo nos permite buscar, mostrar, filtrar, agrupar, … distintos valores de campos sino que también nos permite reconfigurar un campo o, incluso, crear nuevos campos recalculados sobre unos ya existentes.

Además de un gran buscador es un potente visualizador de datos, que nos permite crear y guardar gráficas de muchos tipos para el análisis de datos como:

- Gráficos de barras
- Gráficos de tarta
- Gráficos de líneas

También nos permite crear:

- Tablas: Permiten generar nuevas tablas y exportarlas a csv.

- Área: Para valores acumulados en el tiempo.
- Mapas: cuando tenemos valores geográficos.
- Redes: Explorar las relaciones entre los distintos campos.

Paneles (DashBoards)

Al tener toda una serie de gráficas y otros elementos guardos que nos sirven para el análisis y con una fuente de entrada de datos constante, el dashboard nos permite definir tableros de control para ver cómo están entrando los datos. Su configuración es muy simple, solo hay que apretar el botón «Add», seleccionar las gráficas y/o búsquedas que ya tenemos grabadas y se mostrarán en el tablero, que podremos configurar a nuestro gusto.

Detección de Anomalías con Machine Learning

En la licencia básica no se dispone de toda la funcionalidad completa de Machine Learning (Véase Anexo I) aunque, a continuación, se detallarán las fucionalidades disponibles para la detección de anomalías que dispone la plataforma ELK mediante la utilización de Machine Learning.

Encuentra anomalías y valores atípicos, pronostica con base en tendencias e identifica áreas de interés en tus datos con el aprendizaje automático de Elastic.

Dispone de herramientas de Data Visualizer y Transforms te ayuda a encontrar los trabajos del "droide" necesarios e identifica los campos en tus datos que trabajarían bien con aprendizaje automático. El aprendizaje automático no supervisado con Elastic ayuda a encontrar patrones en los datos. Usa los modelos de series de tiempo para detectar anomalías en los datos actuales y prevé tendencias con base en datos históricos.

Dentro del widget de la tabla Anomalías que se muestra en las páginas Hosts, Red y Detalles asociados, o incluso es posible limitarse al rango de fechas específico de una anomalía desde los detalles de "Max Anomaly Score" en la descripción general de las páginas "Host" e "IP Details". Cada una de las interfaces también ofrecen la capacidad de arrastrar y soltar detalles de la anomalía en el "Timeline".

Para la detección de anomalías es necesario la definición de trabajos que se pueden definir, visualizar, lanzar o parar desde la interfaz de usuario aunque la herramienta de SIEM ya dispone de trabajos predefinidos para la detección de anomalías siempre que utilicemos cualquiera de los agentes Beats que hemos explicado con anterioridad. Para ello basta con configurar el entorno con los índices apropiados correspondientes a cada agente (auditbeat-*, filebeat-*, packetbeat-* o winlogbeat-*) y su integración será directa con los trabajos predefinidos, entre los que habría que destacar:

rare_process_by_host_windows_ecs / rare_process_by_host_linux_ecs

Identifica procesos raros que generalmente no se ejecutan en hosts de Windows / Linux, lo que puede indicar la ejecución de servicios no autorizados, malware o mecanismos de persistencia.

Los procesos se consideran raros cuando solo se ejecutan ocasionalmente en comparación con otros procesos que se ejecutan en el host.

Son necesarios los agentes: Auditbeat (Linux) y/o Winlogbeat (Windows).

windows_anomalous_network_activity_ecs / linux_anomalous_network_activity_ecs

Identifica procesos de Windows / Linux que generalmente no usan la red pero que tienen actividad inesperada en la red, lo que podría indicar actividad de comando y control, movimiento lateral, persistencia o exfiltración de datos.

Un proceso con actividad de red inusual puede denotar explotación o inyección de proceso, donde el proceso se utiliza para ejecutar mecanismos de persistencia que permiten a un actor malicioso el acceso remoto o el control del host, la filtración de datos y la ejecución de aplicaciones de red no autorizadas.

Son necesarios los agentes: Auditbeat (Linux) y/o Winlogbeat (Windows).

windows_anomalous_process_all_hosts_ecs/ linux_anomalous_process_all_hosts_ecs

Busca procesos raros que se ejecutan en varios hosts de Windows / Linux en una red completa.

Esto reduce la detección de falsos positivos ya que los procesos de mantenimiento automatizado a menudo solo se ejecutan ocasionalmente en una sola máquina, pero son comunes a todos o muchos hosts en una red.

Son necesarios los agentes: Auditbeat (Linux) y/o Winlogbeat (Windows).

windows_anomalous_process_creation / windows_anomalous_script

Identifica relaciones inusuales de proceso padre / hijo que podrían indicar la ejecución de malware o mecanismos de persistencia.

Los scripts maliciosos a menudo recurren a otras aplicaciones y procesos como parte de su carga útil de explotación. Por ejemplo, cuando un documento de Office malicioso ejecuta scripts como parte de una carga útil de explotación, Excel o Word pueden iniciar un proceso de intérprete de scripts, que, a su vez, ejecuta un script que descarga y ejecuta malware. Otro escenario común es que Outlook ejecuta un proceso inusual cuando se descarga malware en un correo electrónico.

El segundo trabajo predefinido, busca scripts de PowerShell con características de datos inusuales, como la ofuscación, que pueden ser una característica de los bloques de texto de scripts de PowerShell maliciosos.

Monitorear e identificar relaciones de proceso anómalas es una excelente manera de detectar malware nuevo y emergente que aún no es reconocido por los antivirus.

Es necesario el agente: Winlogbeat (Windows).

linux_anomalous_network_port_activity_ecs / linux_anomalous_network_service

Identifica actividad de puerto de destino inusual que podría indicar “command-and-control”, mecanismo de persistencia o actividad de exfiltración de datos.

La actividad del puerto de destino que se usa con poca frecuencia es generalmente inusual en las flotas de Linux y puede indicar acceso no autorizado o actividad de alguna amenaza.

El segundo trabajo predefinido, busca puertos de escucha inusuales que puedan indicar la ejecución de servicios no autorizados, puertas traseras o mecanismos de persistencia.

Es necesario el agente: Auditbeat (Linux).

linux_anomalous_network_url_activity_ecs

Busca solicitudes de URL web inusuales de los hosts, lo que podría indicar la entrega y ejecución de malware.

Wget y cURL son comúnmente utilizados por los programas de Linux para descargar código y datos. La mayoría de las veces, su uso es completamente normal. Generalmente, debido a que usan una lista de URL, se descargan repetidamente desde las mismas ubicaciones. Sin embargo, Wget y cURL a veces se usan para entregar cargas útiles de explotación de Linux, y los actores de amenazas usan estas herramientas para descargar software y código adicional. Por estas razones, las URL inusuales pueden indicar descargas no autorizadas o actividad de amenaza.

Es necesario el agente: Auditbeat (Linux).

suspicious_login_activity_ecs

Identifica un inusual número de intentos de autenticación.

Son necesarios los agentes: Auditbeat (Linux) y/o Winlogbeat (Windows).

Los trabajos de Machine Learning anteriores analizan dos semanas de datos históricos antes del momento en que están habilitados. Después de habilitar los trabajos, analizan continuamente los datos entrantes. Cuando los trabajos se detienen y reinician dentro del plazo de dos semanas, los datos analizados previamente no se procesan nuevamente.

SIEM Elastic

Existen multitud de herramientas SIEM disponibles en el mercado, tanto comerciales como de Opensource. Una vez que ya nos hemos decantado por la solución ELK Stack, este paquete dispone de un SIEM (podemos ver las características completas incluidas en la suscripción básica gratuita en el Anexo I) con probada capacidad de integración y potencialidad y que será nuestro elegido para la implementación e integración con los datos recopilados en las fases anteriores.

Proporciona integraciones de red y datos de host, analíticas compartibles con base en Elastic Common Schema (ECS) y la capacidad de explorar tus datos de seguridad con la app SIEM en Kibana. Con integraciones de Beats permite explotar:

- ✓ Eventos de auditoría: Utiliza Auditbeat para monitorear el sistema y los detalles de integridad de archivos y envía a Elasticsearch para analizarlos en la app SIEM.

- ✓ Logs de autenticación: Investiga intentos de inicio de sesión y actividades relacionadas con los datos de autenticación recopilados por Auditbeat y los módulos de sistema de Filebeat.

- ✓ Tráfico de DNS: Visualiza lo que sucede con los datos de DNS recopilados por Packetbeat: patrones de acceso de usuario, actividad de dominio, tendencias de búsqueda y más.

- ✓ NetFlow: Establece una visión amplia de tu entorno a partir del análisis de los datos de flujo, recolectados y analizados por el módulo de Netflow de Filebeat.

- ✓

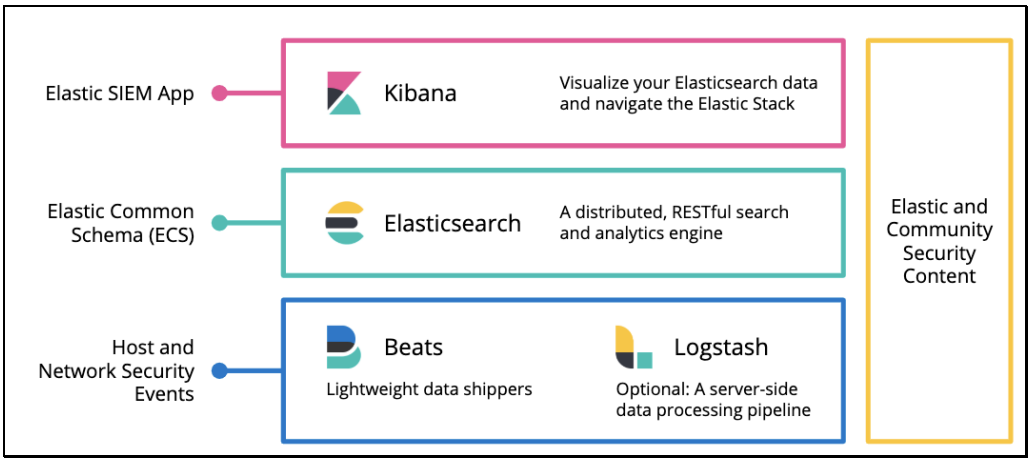


Figura 3 - Componentes necesarios para SIEM (Fuente Elastic)

En la licencia básica, no se dispone de toda la funcionalidad completa de SIEM Elastic (Véase Anexo I).

Interfaz Usuario SIEM

La aplicación SIEM es un espacio de trabajo altamente interactivo muy útil para analistas de seguridad. Diseñada para ser detectable, clicable, arrastrable y soltable, expandible y plegable, redimensionable, movible, etc. Se empezaría con una descripción general y posteriormente se puede usar la IU interactiva para profundizar en áreas de interés.

La **Vista Hosts** proporciona métricas clave con respecto a eventos de seguridad relacionados con el host, y tablas de datos y widgets que permiten interactuar con el Visor de eventos de la línea de tiempo. Se puede obtener información más profunda y arrastrar y soltar elementos de interés desde las tablas de vista de Hosts a la Línea de tiempo para una mayor investigación.

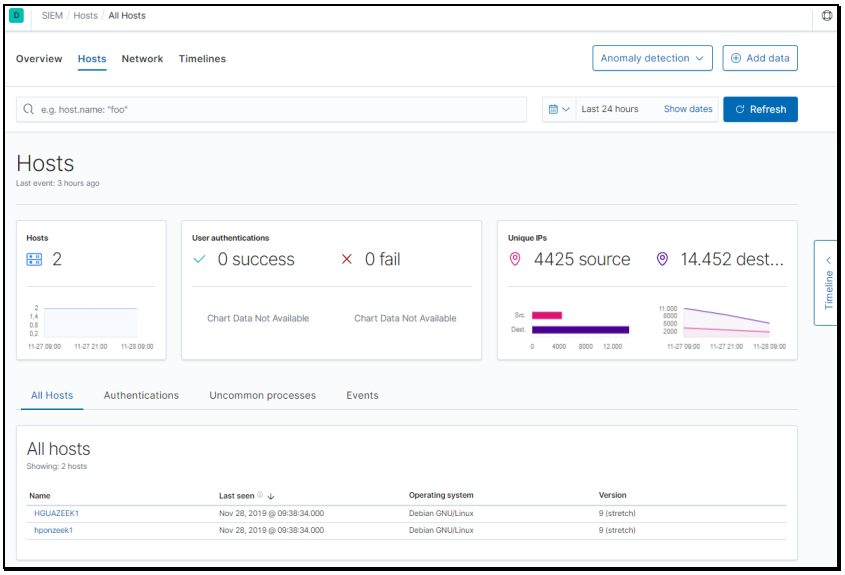


Figura 4 - SIEM Vista Hosts

La **Vista Network** proporciona métricas clave de actividad de red, facilita el enriquecimiento del tiempo de investigación y proporciona tablas de eventos de red que permiten la interacción con la Línea de tiempo. Se puede obtener información más profunda y arrastrar y soltar elementos de interés desde la vista de red a la línea de tiempo para una mayor investigación.

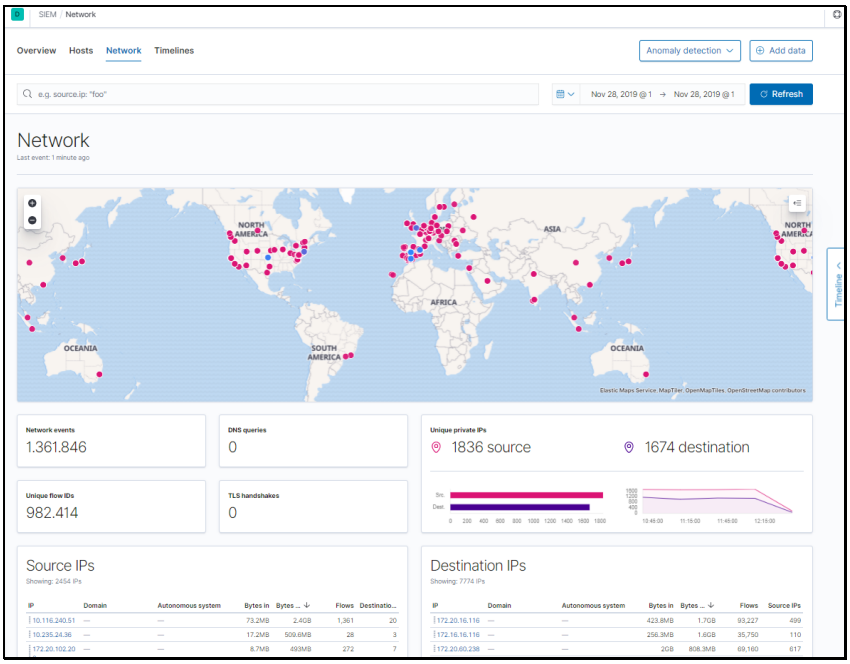


Figura 5 - SIEM Vista Network

El **Timeline** o línea de tiempo es el espacio de trabajo para la búsqueda de amenazas y las investigaciones de alerta. Se pueden arrastrar objetos de interés al Visor de eventos de la línea de tiempo para crear exactamente el filtro de consulta que se necesita. También se puede arrastrar elementos desde los widgets de la tabla dentro de las páginas Hosts y Network, o incluso desde la propia línea de tiempo.

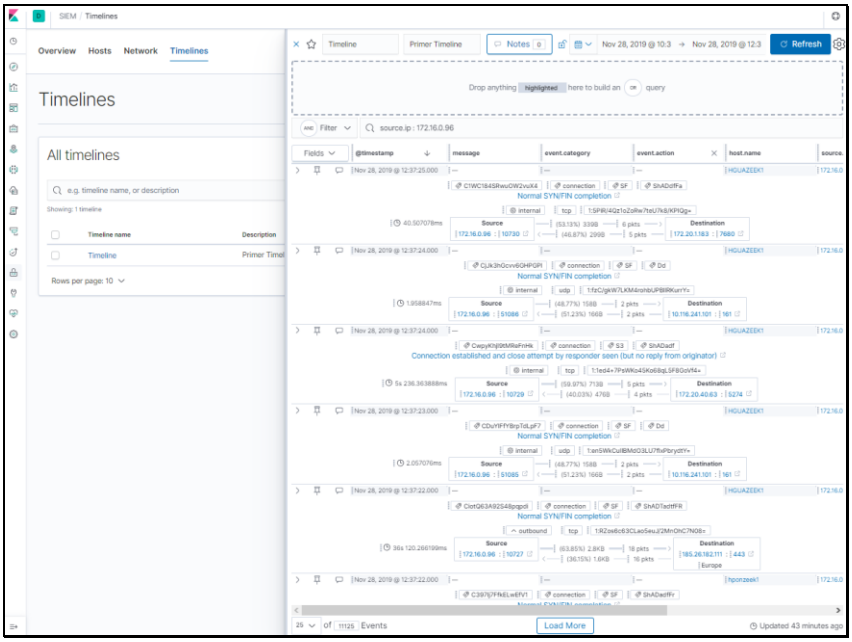


Figura 6 - SIEM Timeline

Si un analista detecta algún comportamiento o actividad sospechosa que justifica una mayor investigación, y hace clic en una URL que enlaza con la aplicación SIEM. Se pueden utilizar las tablas, los widgets y las capacidades de filtrado y búsqueda en la aplicación SIEM para llegar hasta el final de la alerta. También se pueden arrastrar elementos de interés a la línea de tiempo para su posterior análisis.

En definitiva, dentro de la línea de tiempo, un analista puede investigar más a fondo (profundizar, buscar y filtrar) y agregar notas y fijar elementos de interés.

Otros componentes ELK Stack

Existen otros componentes en la plataforma ELK que pueden resultar interesantes y entre los que destacan:

Alerting

Permite enviar notificaciones como que el uso de la CPU está aumentando inesperadamente, el tiempo de respuesta de las aplicaciones está incrementando, los errores 503 se están disparando, la tasa de indexación de Elasticsearch ha disminuido considerablemente, etc.

Como característica importante, permite combinar las alertas con características de Machine Learning sin supervisión para encontrar comportamientos inusuales ya que usa las puntuaciones de anomalías en el marco de trabajo de alertas para recibir notificaciones cuando el barco esté fuera de curso.

Se dispone de un historial completo de todas las ejecuciones de alertas indexado en Elasticsearch para facilitar el rastreo y la visualización en Kibana.

Canvas

Permite convertir los datos en dashboards únicos y dinámicos con los logos, los colores y los elementos de diseño que queramos, incluso con marca propia. Tanto para logs de

infraestructura, eventos de seguridad, métricas de aplicación o datos de un proyecto propio, Elastic Canvas proporciona un control creativo.

4.6. Seguridad del Paciente asociada al uso de las TIC

En la medicina el uso de las TIC va desde el diagnóstico de enfermedades hasta el monitoreo de los signos vitales de un paciente en un nivel operatorio o postoperatorio.

Una vez identificadas y decididas las herramientas que nos ayudarán a la detección de anomalías, es necesario documentar el concepto de Seguridad del Paciente y los beneficios que el uso de las TIC nos proporciona. De la misma forma, hay unos riesgos asociados a este uso, entre ellos, los debidos a un mal funcionamiento accidental o intencionado ocasionados por una infección por malware, un error de programación u otro tipo de anomalía.

Concepto de Seguridad del Paciente

Según la Organización Mundial de la Salud (OMS), la seguridad del paciente es un principio fundamental de la atención sanitaria. Hay un cierto grado de peligrosidad inherente a cada paso del proceso de atención de salud. Los eventos adversos pueden estar en relación con problemas de la práctica clínica, de los productos, de los procedimientos o del sistema. La mejora de la seguridad del paciente requiere por parte de todo el sistema un esfuerzo complejo que abarca una amplia gama de acciones dirigidas hacia la mejora del desempeño; la gestión de la seguridad y los riesgos ambientales, incluido el control de las infecciones; el uso seguro de los medicamentos, y la seguridad de los equipos, de la práctica clínica y del entorno en el que se presta la atención sanitaria.

La OMS también indica que la atención sanitaria no debería causar daños a nadie. Sin embargo, cada año millones de pacientes en todo el mundo sufren daños debido a una atención sanitaria poco segura. En los países de ingresos altos, se calcula que uno de cada 10 pacientes sufre daños mientras recibe atención hospitalaria. En el ámbito mundial, hasta cuatro de cada 10 pacientes sufren daños en la atención primaria de salud o los servicios ambulatorios. Los errores más perjudiciales son los relacionados con el diagnóstico y la prescripción y uso de medicamentos. Una de las claves para ofrecer una atención más segura es la colaboración más activa de los pacientes.

Beneficios del uso de las TIC en la Seguridad del Paciente

Las TIC se clasifican según la función que atienden (telemedicina, teleasistencia, organización y procesos, integración y digitalización de la información, movilidad, etc.) y según el tipo de intervención que realizan (atención primaria, atención especializada, salud mental, salud pública…).

Los servicios sanitarios se apoyan cada vez más en el uso de las Tecnologías de la Información Sanitarias (TIS) y resulta difícil imaginar a un sistema sanitario funcionando sin herramientas tales como la Historia Clínica Electrónica, el Sistema de Cuidados de Enfermería, los sistemas de información de laboratorios, de farmacia o los

sistemas PACS (Picture Archiving and Communication System) para el diagnóstico por imagen.

Los beneficios que aporta el uso de las TIC en el sector sanitario son incontables, al igual que las aportaciones relacionadas con la seguridad del paciente. Los sistemas pueden, por ejemplo, alertar al profesional sobre las alergias de un paciente impidiendo que se le prescriba y administre un medicamento que podría ser potencialmente peligroso.

En el ámbito de seguridad del paciente, Alotaibi y Federico han realizado la revisión bibliográfica "The impact of health information technology on patient safety" [ALO17] en la que se observa cierta evidencia de que la implementación de la historia clínica electrónica reduce los errores médicos y mejora la seguridad del paciente. La entrada de orden médica computarizada y el apoyo de decisión clínica son probablemente dos de las TIC más beneficiosas para mejorar la seguridad del paciente. Además, el sistema automático y dispensación de medicamentos y el sistema de gestión de datos del paciente parecen mejorar la seguridad del paciente en las unidades de curas intensivas.

Esta revisión puede ser útil a médicos y responsables de políticas sanitarias a la hora de tomar decisiones basadas en la evidencia sobre la adquisición y la implementación de estas TIC, para mejorar la seguridad del paciente.

En la siguiente tabla se ejemplifica la evidencia sobre diversas TIC y su impacto en la seguridad de los pacientes.

Tipos de TIC	Impacto en la seguridad del paciente
Entrada de orden médica computarizada (COPE)	Reducción de la tasa de errores de medicación (sólo observado cuando se integra con Apoyo de Decisión Clínica)
Apoyo de decisión clínica (CDS)	Mejora del proceso de adherencia, prescripción de medicamentos, vacunación, demandes de laboratorio y resultados clínicos.
Herramientas electrónicas de gestión del paciente	Mejora en los procesos de trasnferencia y menos omisiones de información crítica del paciente. Poca evidencia de reducción de errores médicos.
Administración de medicación por código de barras (BCMA)	Reducción en los errores de medicación y en las reacciones adversas. Reducción en los errores de etiquetaje de los especímenes de laboratorio.
Bombeo inteligente	Insuficiente evidencia en la reducción de errores de medicación. Reducción de errores en la programación del bombeo.
Sistemas de gestión de datos del paciente (PDMS)	Reducción del tiempo de control, más tiempo empleado en el trato directo con el paciente y una reducción de errores.
Sistema automático de dispensación de medicamentos (ADC)	Reducción de errores de medicación en las unidades de cuidados intensivos.
Detector de elementos quirúrgicos retenidos	Ninguna reducción significativa en la tasa de elementos quirúrgicos retenidos.
Portales de pacientes	Mayor cumplimiento de los servicios de prevención médica. Reducción de la frecuencia de ataques de asma. Mejora de la adherencia de los pacientes a la medicación, conciencia de la enfermedad, autogestión de la enfermedad y satisfacción del paciente. Ninguna evidencia en la mejora de la seguridad del paciente.
Telemedicina – visitas virtuales	Tan efectiva como las visitas cara a cara en cuanto a los resultados clínicos. Ninguna evidencia en cuanto a seguridad del paciente.

Telemedicina – telemonitorización	Mejora en los resultados clínicos para pacientes con ciertas enfermedades crónicas como la hipertensión. Ninguna evidencia en cuanto a resultados en seguridad del paciente.
Informe electrónico de incidentes	Incremento significativo en la frecuencia de informes sobre eventos adversos.
Registros médicos electrónicos (EMR)	Mejora en la adherencia a las guías. Reducción de errores médicos. Reducción de reacciones adversas a los medicamentos. Sin impacto significativo en la mortalidad.

Fuente: The impact of health information technology on patient safety

Riesgos del uso de las TIC en la Seguridad del Paciente

En general, los objetivos de la seguridad de la información son mantener la confidencialidad, integridad y disponibilidad de la información.

Seguridad es sinónimo de confianza y se debe hacer todo lo posible para evitar que ésta se rompa o al menos, si se rompe, recuperarla lo antes posible. Esto es extensible tanto a la seguridad clínica como a la seguridad de la información, que deben garantizarse al máximo.

En su informe 11 del año 2018, la Sociedad Española de Informática de la Salud (SEIS), se alertaba de que la incorporación de las Tecnologías de la Tecnologías de la Información y la Comunicación (TIC) a la atención sanitaria es un hecho consolidado desde hace mucho tiempo, y se ha llevado a cabo con el objetivo de mejorar la eficacia, eficiencia y efectividad del sistema de salud. Sin embargo, y a pesar de su potencial, las TIC pueden suponer también la introducción de riesgos para la seguridad del paciente. La gestión de estos riesgos debe comenzar con un análisis pormenorizado de los mismos, para definir sus principales parámetros, evaluar sus posibles consecuencias y, finalmente, decidir la estrategia más adecuada para afrontarlos.

A continuación, vamos a intentar asociar cómo puede afectar el uso de la tecnología a los errores más perjudiciales para la seguridad del paciente indicados anteriormente según la OMS:

Errores en el Diagnóstico

Por ejemplo, el empleo de las tecnologías de la información y la comunicación (TIC) forman parte importante, si no ineludible, en el manejo de este proceso, al punto tal que no se puede concebir un laboratorio moderno sin sistema de información. Sin embargo, es importante tomar en cuenta que este aporte genera potencialmente nuevas posibilidades de error, que indefectiblemente deben ser consideradas y controladas.

Se han documentado casos de la empresa Carestream Health, Inc., dedicada a las soluciones de software de imagen médica de radiodiagnóstico, donde es posible la modificación de imágenes radiológicas mediante distintas técnicas de inteligencia artificial que pueden derivar en un diagnóstico distinto del real; esto es, puede parecer la exitencia un tumor donde realmente no lo había y viceversa.

Entonces, asumiendo que existen multitud de equipos que intervienen en la ayuda para el diagnóstico, nosotros nos centraremos en aquellos que están conectados a la red y que serán los que entrarían dentro del ámbito de aplicación de este trabajo. Esto es:

- Equipos de radiodiagnóstico (TAC, Radiografía Convencional, Resonancia Magnética Nuclear, Ecógrafos, …).

- Equipos de análisis clínicos (Coagulómetros, Orina, Hematología, …)

- Equipos de monitorización (Electrocardiogramas, Tensión arterial, Saturación, Pulso, …)

Todos estos equipos se encuentran conectados en rangos de red específicos cuya monitorización por el sistema puede realizarse de forma efectiva. De esta forma, podríamos identificar comportamientos anómalos de estos equipos, conexiones no autorizadas utilizando metodologías similares desde otros equipos que pudiesen poner en peligro la seguridad de la información o incluso ser un riesgo para la seguridad del paciente si este comportamiento anómalo proporciona valores que no se corresponden con los resultados reales.

Errores en la Prescripción de Medicamentos

La seguridad de la información, tanto en su vertiente técnica como normativa, debe ser siempre un elemento de suma y no de resta en la seguridad clínica. Como algunos aspectos de la seguridad de la información pueden ser contrarios a la seguridad de los pacientes, siempre se deben ponderar los distintos aspectos que intervienen en su ejecución, para lograr un equilibrio que por un lado asegure la seguridad de la información y por otro no ponga en peligro la seguridad de los pacientes.

Por ejemplo, proteger con controles excesivos el acceso a la información clínica (confidencialidad, un componente de la seguridad de la información) puede ser perjudicial para la seguridad del paciente, si se dificulta tanto ese acceso que los clínicos desisten de consultar información que puede ser trascendental para la asistencia.

En la prescripción de los medicamentos, en nuestro caso, intervendrían:

- La Historia Digital del Paciente, donde incluimos el software específico de Farmacia para tal fin.

- Equipos de dispensación e imputación de unidosis conectados a la Historia Digital.

Al igual que para el caso anterior, los equipos se encuentran configurados en rangos de red determinados que nos ayudará a una monitorización activa para la detección de distintas anomalías. Por ello, sería necesario la identificación de aquellos equipos que establezcan una conexión la base de datos de forma no autorizada y que pudiese ocasionar una modificación de parámetros, dosis u otra información que pueda llevar a una medicación distinta a la establecida originalmente por el facultativo.

Errores en el Uso del Medicamento

En la clasificación realizada por la American Society of Health-System Pharmacists en 1993 que basadas en 12 categorías diferentes de errores, éstos pueden presentarse debido a fallos activas por la práctica de los diferentes profesionales involucrados o a fallas latentes en el sistema de utilización de los medicamentos, bien sea en la prescripción médica, en la dispensación por farmacia o en la administración y monitorización que realiza el profesional de enfermería.

En nuestro caso, no se dispone de ningún equipo electromédico conectado a nuestra red que intervenga en la administración directa de los medicamentos. De esta forma, no entrará dentro del ámbito de estudio de este trabajo.

Otros riesgos

Es evidente que, a parte de los descritos anteriormente, existen otros muchos riesgos derivados del uso de las nuevas tecnologías de la información tanto para la seguridad del paciente como para la seguridad misma de la información, aunque no entran dentro del ámbito de este trabajo.

Lo que sí que hay que destacar es que para reducir cualquier riesgo es necesario incorporar a la seguridad del paciente a los objetivos de todo proyecto de desarrollo de un sistema TIC sanitario, y tenerla en cuenta en todos los procedimientos de trabajo y de gestión del proyecto. Existen algunas recomendaciones y buenas prácticas que pueden ser de ayuda para este tipo de actuaciones, como por ejemplo las guías SAFER.

5. Implementación

5.1. Aprovisionamiento de recursos hardware y software

Como hemos indicado en anteriores capítulos, para la implementación de la solución hemos recuperado 2 PCs que estaban sin uso debido a una renovación de equipamiento informático y que serán los que destinaremos para la instalación de Debian+Zeek (Bro IDS)+Filebeat. Cada PC se instalará físicamente en los CPDs de 2 centros distintos.

Se ha procedido a la descarga e instalación del siguiente software:

- Debian 9 → https://cdimage.debian.org/cdimage/archive/9.9.0/amd64/iso-dvd/debian-9.9.0-amd64-DVD-1.iso
- Zeek 3.0.0 → https://www.zeek.org/downloads/zeek-3.0.0.tar.gz
- Filebeat 7.4.1 → https://artifacts.elastic.co/downloads/beats/filebeat/filebeat-7.4.1-linux-x86_64.tar.gz

Se ha creado una máquina Windows 2012 Server con 16 GB de RAM, 8 CPUs Virtuales, 1 HD de 20 GB para Sistema Operativo, 1 HD de 200 GB para Elasticsearch y 1 HD de 10 GB para el resto de paquetes de ELK Stack.

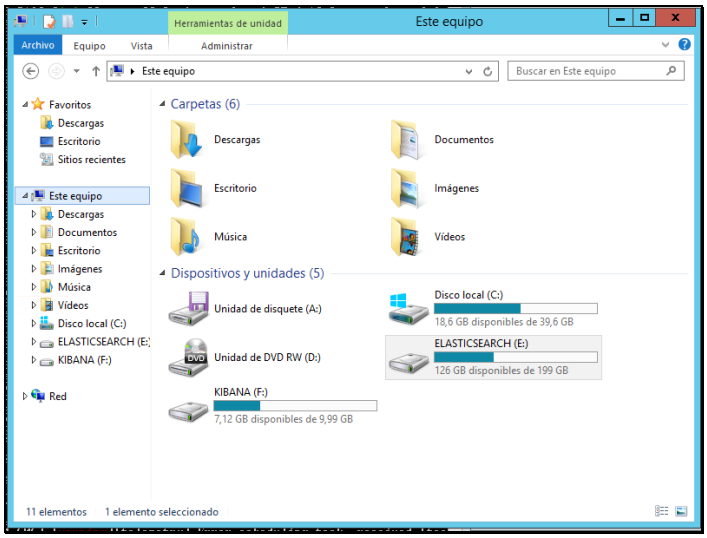


Figura 7 - Almacenamiento ELK Stack

Se ha descargado el software:

- Elasticsearch 7.4.0 → https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.4.1-windows-x86_64.zip
- Kibana 7.4.1 → https://artifacts.elastic.co/downloads/kibana/kibana-7.4.1-windows-x86_64.zip

5.2. Instalación y configuración ELK Stack

Para realizar la instalación, vamos a seguir las indicaciones de la guía de instalación de la web oficial:

- https://www.elastic.co/guide/en/elastic-stack/current/installing-elastic-stack.html

En esta guía, se indica que la instalación de los productos, si queremos hacer la instalación completa, debe ser en este orden para que no haya problemas de componentes faltantes. Además, todos los componentes deberían estar en la misma versión (en nuestro caso la última disponible es la 7.4.1):

Instalación de Elasticsearch

Opción 1: Básica

El paquete para Windows es un fichero con extensión .zip y que contiene todos los componentes necesarios para su ejecución, incluyendo una máquina virtual java. Lo descomprimimos en la carpeta "e:\elasticsearch7.4.1\".

La versión que hemos descargado tiene incluidas todas las características disponibles aunque el alcance de este trabajo sólo tratará sobre aquellas que son gratuitas (ver ANEXO I). Para el desarrollo de este proyecto se pretende que el coste de licenciamiento y de hardware sea 0 o no sea necesaria ninguna inversión inicial.

Tenemos que modificar el fichero "elastic.yml" para que acepte conexiones desde otro equipo, por defecto sólo acepta desde localhost, ya que la carga de datos se realizará desde los equipos que vamos a instalar Zeek (Bro IDS).

Instalamos Elasticsearch como servicio y lo arrancamos:

- e:\elasticsearch-7.4.1\bin\elasticsearch-service.bat install

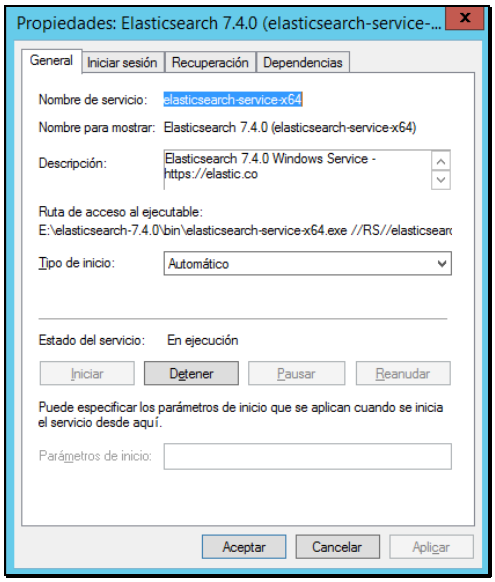


Figura 8 - ElasticSearch como Servicio

Opción 2: Instalación mediante fichero ejecutable MSI.

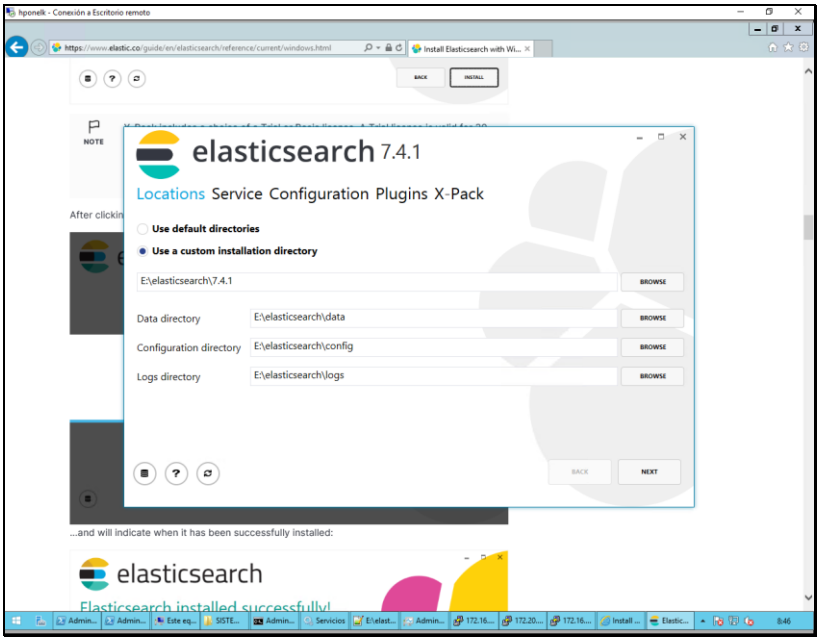


Figura 9 - Instalación Elastic Search Licencia

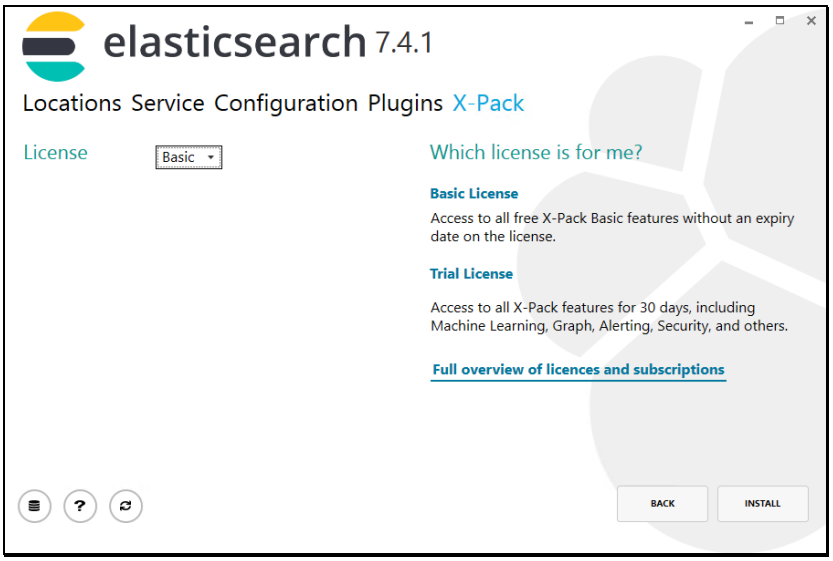


Figura 10 - Instalación Elastic Search Licencia

Instalación de Kibana

El paquete para Windows es un fichero con extensión .zip y que contiene todos los componentes necesarios para su ejecución. Lo descomprimimos en la carpeta "f:\kibana-7-4-1".

La versión que hemos descargado tiene incluidas todas las características disponibles aunque el alcance de este trabajo sólo tratará sobre aquellas que son gratuitas (ver ANEXO I).

Tenemos que modificar el fichero "kibana.yml" para que acepte conexiones desde otro equipo, por defecto sólo acepta desde localhost, ya que la carga de datos se realizará

desde los equipos que vamos a instalar Bro IDS. No es necesario modificar la configuración sobre Elasticsearch ya que, en nuestro caso, están instaladas en la misma máquina (localhost).

Por último, arrancamos el fichero "HOME_KIBANA\bin\kibana.bat"

5.3. Instalación y configuración ZEEK (IDS BRO)

Es importante destacar que el funcionamiento de Zeek se basa en esnifar tráfico de red forzando a una o varias interfaces de red de nuestro equipo a trabajar en modo promiscuo.

Antes de instalar el software es necesario determinar la ubicación física en la que vamos a poner a trabajar los distintos equipos. La ubicación ideal sería aquella en la que más tráfico de red pueda pasar y ésta sería en el CPD que es donde están los switches principales, los servidores y los armarios de comunicaciones con el exterior. Una ubicación incorrecta puede ocasionar que el tráfico que capturamos no sea relevante para la detección de amenazas o de anomalías.

Una vez seleccionada la ubicación, tendríamos que decidir qué topología de red utilizar para capturar el mayor tráfico posible con la menor interferencia en la red y en el rendimiento de las aplicaciones. Tenemos 2 opciones:

1. Configurarlo como router y que todo el tráfico pase por el equipo y que luego se reenvíe a nuestro Firewall/Router para salir al exterior.

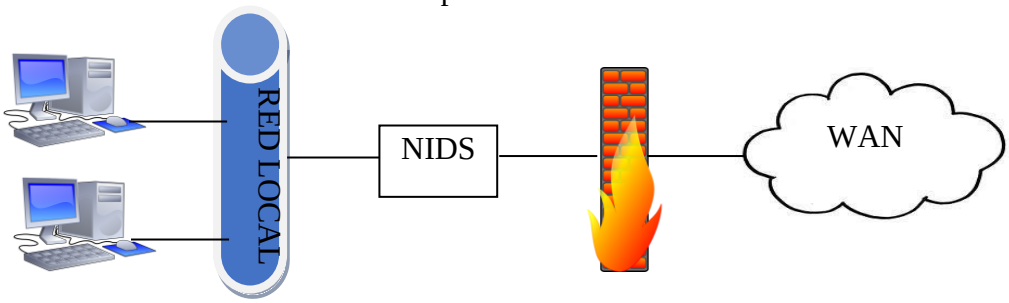


Figura 11 - NIDS en línea

2. Hacer "Port Mirroring" de las bocas donde están conectadas nuestro Router/Firewall y que sean monitorizadas por la boca donde conectamos el equipo que vamos a instalar. Para esto, es necesario que el Switch sea "gestionado", no vale un hub o switch de tipo "no gestionado".

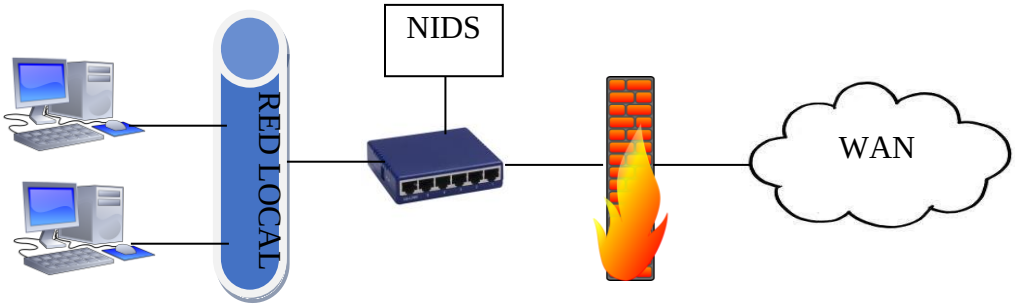


Figura 12 - NIDS con Port Mirroring

Dada la baja capacidad de los equipos que vamos a montar, implementar la primera alternativa puede ocasionar una importante lentitud en nuestra red. De esta forma, procedemos a implantar la segunda alternativa en los dos centros donde vamos a capturar el tráfico de red ya que disponemos de switches gestionados de última generación y es la que no ocasiona interferencias en el trabajo diario en la comunicación de los distintos sistemas en producción.

A continuación, realizamos la instalación en ambos equipos, que disponen de la versión de linux Debian 9 previamente instalada en pasos anteriores, según la guía de instalación oficial (https://docs.zeek.org/en/stable/install/install.html), los comandos se lanzarán como root o utilizando "sudo":

- Instalamos pre-requisitos

apt-get install cmake make gcc g++ flex bison libpcap-dev libssl-dev python-dev swig zlib1g-dev

- Instalamos Zeek

sudo git clone --recursive https://github.com/zeek/zeek

cd /root/zeek

./configure

make

make install

export PATH=/usr/local/zeek/bin:$PATH

apt-get install gdb

- Instalamos opciones de geolocalización para identificar el origen o destino de las IPs fuera de la red local:

apt-get install libmaxminddb-dev

- Configuramos Zeek para que el formato de salida de los logs sea JSON en lugar de ASCII. Para ello:
  - En 'ZEEK_HOME/share/zeek/base/frameworks/logging/writers/ascii.zeek' cambiamos "const use_json = F &redef;" por "const use_json = T &redef;"
  - Añadimos en ''ZEEK_HOME/ share/zeek/site/local.zeek" la siguiente línea "@load policy/tuning/json-logs.zeek

- Arrancamos Zeek:

ZEEK_HOME/bin/zeekctl

zeekctl> deploy

Una vez arrancado Zeek comienza a esnifar todo el tráfico de red de las tomas que hemos configurado en el switch para monitorizar. El tipo de tráfico de red detectado y que podremos integrar con ELK Stack es el que podemos visualizar en la siguiente figura:

/usr/local/zeek/spool/zeek/

Nombre	Tamaño	Modificado	Permisos
..		20/11/2019 11:50:43	rwxr-xr-x
capture_loss.log	8 KB	18/11/2019 14:20:47	rw-r--r--
conn.log	2.159.168 KB	18/11/2019 15:08:57	rw-r--r--
dce_rpc.log	164.704 KB	18/11/2019 14:41:48	rw-r--r--
dhcp.log	1.271 KB	18/11/2019 14:26:47	rw-r--r--
dns.log	687.701 KB	18/11/2019 14:41:48	rw-r--r--
dpd.log	4.169 KB	18/11/2019 14:41:47	rw-r--r--
files.log	1.082.308 KB	18/11/2019 14:41:48	rw-r--r--
ftp.log	515 KB	18/11/2019 14:03:07	rw-r--r--
http.log	887.088 KB	18/11/2019 14:41:48	rw-r--r--
kerberos.log	57.082 KB	18/11/2019 14:41:48	rw-r--r--
known_certs.log	497 KB	18/11/2019 14:41:48	rw-r--r--
known_hosts.log	73 KB	18/11/2019 14:41:44	rw-r--r--
known_services.log	2.647 KB	18/11/2019 14:26:47	rw-r--r--
mysql.log	1.398 KB	18/11/2019 14:41:45	rw-r--r--
notice.log	1.167 KB	18/11/2019 14:25:47	rw-r--r--
ntlm.log	160.925 KB	18/11/2019 14:41:48	rw-r--r--
ntp.log	22.584 KB	18/11/2019 14:41:45	rw-r--r--
pe.log	115 KB	18/11/2019 14:20:46	rw-r--r--
rdp.log	41 KB	18/11/2019 14:25:46	rw-r--r--
smb_files.log	211.934 KB	18/11/2019 14:41:48	rw-r--r--
smb_mapping.log	26.932 KB	18/11/2019 14:41:48	rw-r--r--
smtp.log	2.765 KB	18/11/2019 14:26:46	rw-r--r--
snmp.log	1.951 KB	18/11/2019 14:26:20	rw-r--r--
socks.log	248 KB	18/11/2019 14:25:49	rw-r--r--
software.log	3.738 KB	18/11/2019 14:41:46	rw-r--r--
ssh.log	267 KB	18/11/2019 14:23:38	rw-r--r--
ssl.log	229.713 KB	18/11/2019 14:41:48	rw-r--r--
stats.log	92 KB	18/11/2019 14:25:47	rw-r--r--
stderr.log	1 KB	18/11/2019 0:06:31	rw-r--r--
stdout.log	1 KB	12/11/2019 9:36:24	rw-r--r--
tunnel.log	176 KB	18/11/2019 14:41:47	rw-r--r--
weird.log	61.256 KB	18/11/2019 14:41:47	rw-r--r--
x509.log	170.573 KB	18/11/2019 14:41:48	rw-r--r--

Figura 13 - Logs tráfico de red detectado por Zeek

5.4. Integración Zeek-ELK Stack

Para cargar la información desde Zeek (Bro IDS), podemos hacerlo utilizando Logstash o, también en las últimas versiones, desde Filebeat, ambas herramientas disponibles en ELK Stack.

Hemos optado por realizarlo con Filebeat que contiene un módulo específico, entre otros, para Zeek. Filebeat es un software ligero para reenviar y centralizar datos de registro. Al instalarlo como agente en los equipos o servidores, Filebeat supervisa los archivos de registro o las ubicaciones que especifique, recopila eventos de registro y los reenvía a Elasticsearch o Logstash para su indexación.

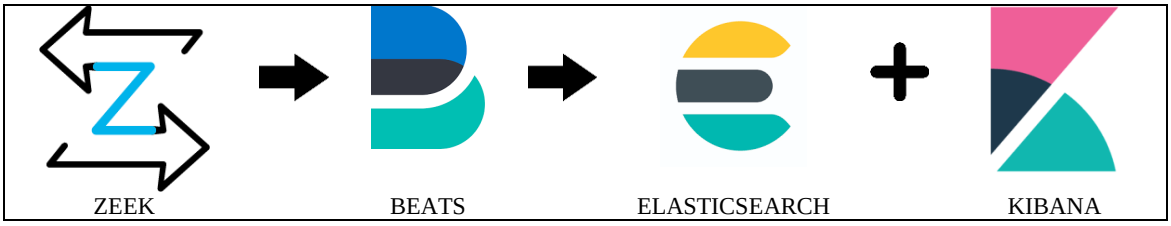


Figura 14 - Integración Zeek & ELK Stack

Hay 2 opciones de configuración para Debian:

Opción 1: Manual

- Descomprimimos el fichero.gz.zip descargado en $FILEBEAT_HOME:
- Modificar el fichero "filebeat.yml" para indicarle las IPs del servidor de Elasticsearh y de Kibana, que en nuestro caso están en la misma ubicación.
- Habilitamos el dashboard de kibana para pasarle cierta información directamente:

$FILEBEAT_HOME/filebeat setup --modules zeek -e -E 'setup.dashboards.enabled=true'

- Habilitamos el módulo "zeek" de Filebeat:

$FILEBEAT_HOME/filebeat modules enable zeek

- Modificamos las rutas de los logs que vamos a capturar, por si no coinciden con la configuración por defecto.
- Arrancamos Filebeat en modo visualización:

$FILEBEAT_HOMEfilebeat -e

Opción 2: Como servicio

- Descargamos el paquete .deb de la página oficial y lo instalamos:

curl -L -O https://artifacts.elastic.co/downloads/beats/filebeat/filebeat-7.4.1-amd64.deb

sudo dpkg -i filebeat-7.4.1-amd64.deb

- Cargamos la plantilla de índices para ELK:

filebeat setup --index-management -E output.logstash.enabled=false -E 'output.elasticsearch.hosts=["elasticsearch_host:9200"]'

- Modificar el fichero "filebeat.yml" para indicarle las IPs del servidor de Elasticsearh y de Kibana, que en nuestro caso están en la misma ubicación.
- Configuramos los dashboards para Kibana.

filebeat setup --dashboards

- Habilitamos el dashboard de kibana para pasarle cierta información directamente:

$FILEBEAT_HOME/filebeat setup --modules zeek -e -E 'setup.dashboards.enabled=true'

- Habilitamos el módulo "zeek" de Filebeat:

filebeat modules enable zeek

- Modificamos las rutas de los logs que vamos a capturar, por si no coinciden con la configuración por defecto.
- Arrancamos Filebeat en modo servicio:

service filebeat start

En ambas instalaciones, es importante definir las rutas exactas donde están nuestros ficheros de logs de Zeek para la transformación e importación a Elastic Search:

```
var:
  - name: paths
    default:
      - /usr/local/zeek/spool/zeek/conn.log
    os.linux:
      - /usr/local/zeek/spool/zeek/conn.log
    os.darwin:
      - /usr/local/var/logs/current/conn.log
  - name: tags
    default: [zeek.connection]
  - name: community_id
    default: true

ingest_pipeline: ingest/pipeline.json
input: config/connection.yml

requires.processors:
- name: geoip
  plugin: ingest-geoip
```

Figura 15- Ubicación Logs Zeek para Filebeat

Filebat nos permite seleccionar qué logs, de todos los que captura Zeek, queremos transformar y enviar. La versión descargada por defecto, al habilitar el módulo Zeek, ya viene preconfiguradas varias transformaciones para cargar (connection, dns, http, files, ssl, notice).

En las pruebas realizadas hemos visto que hay logs que diariamente alcanzan varios gigas de tamaño con lo que, debido a esta gran cantidad de información y que en el log de conexiones (connection) ya disponemos de la mayoría de información contenida en los otros logs, solamente dejamos habilitada la opción de "connection" en la configuración de nuestro "Filebeat":

```
# Module: zeek
# Docs: https://www.elastic.co/guide/en/beats/filebeat/7.4/filebeat-module-zeek.html

- module: zeek
  # All logs
  connection:
    enabled: true
  dns:
    enabled: false
  http:
    enabled: false
  files:
    enabled: false
  ssl:
    enabled: false
  notice:
    enabled: false

    # Set custom paths for the log files. If left empty,
    # Filebeat will choose the paths depending on your OS.
    #var.paths:
```

Figura 16 - Transformación Logs Filebeat

Como hemos indicado, Zeek es capaz, por defecto, de capturar más información de la que realmente vamos a usar en este trabajo incluso procesarla mediante la utilización de scripts. Debido a las características especiales de tiempo para la realización de un Trabajo de Fin de Máster sólo nos centraremos en el tráfico "connection".

Para avanzar más, ya hay algunos proyectos interesantes que han desarrollado más transformaciones para enviarlas a Elastic Search. Un ejemplo claro lo podemos ver en

https://github.com/elastic/beats/pull/12812 (“[WIP] [Filebeat] Complete Zeek module”):

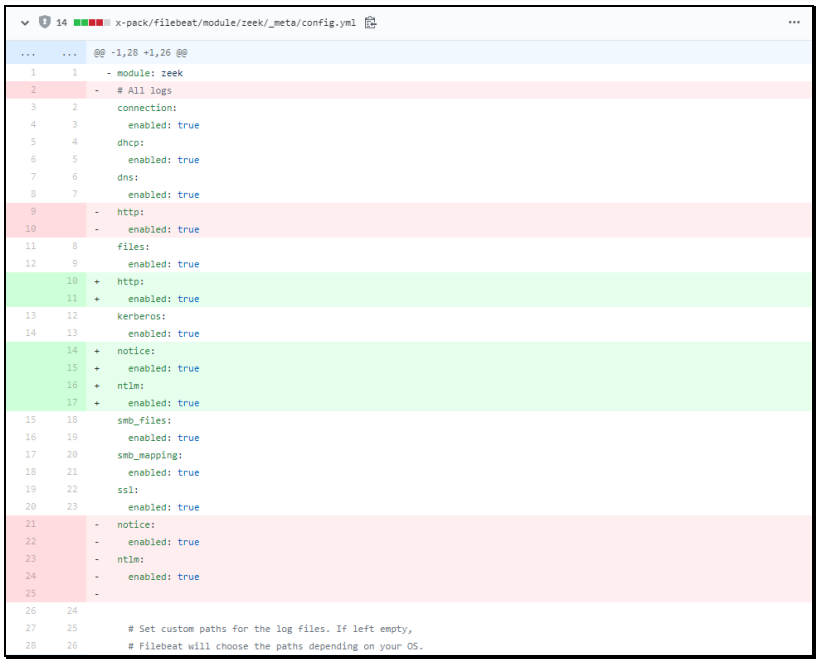


Figura 17 - Módulos Integración Zeek-Filebeat GitHub

Continuamos con la comprobación de la llegada correcta de los datos a Elasticsearch. Para ello, abrimos Kibana desde el entorno web (http://localhost/5601) y vemos que hay disponible un índice en Elasticsearch para trasladarlo a Kibana y procedemos a su creación en Kibana, una vez que seleccionamos el filtro de tiempo por el que queremos que se ordenen nuestros logs en Kibana:

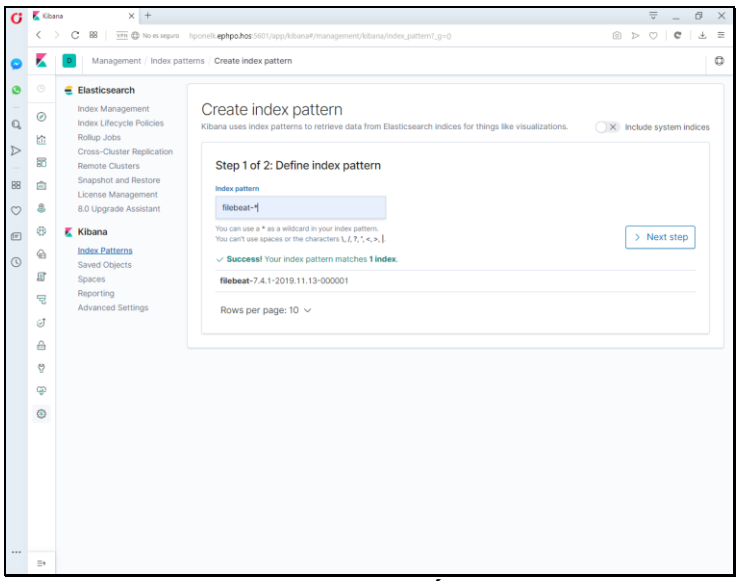


Figura 18 - Creación de Índice Paso 1

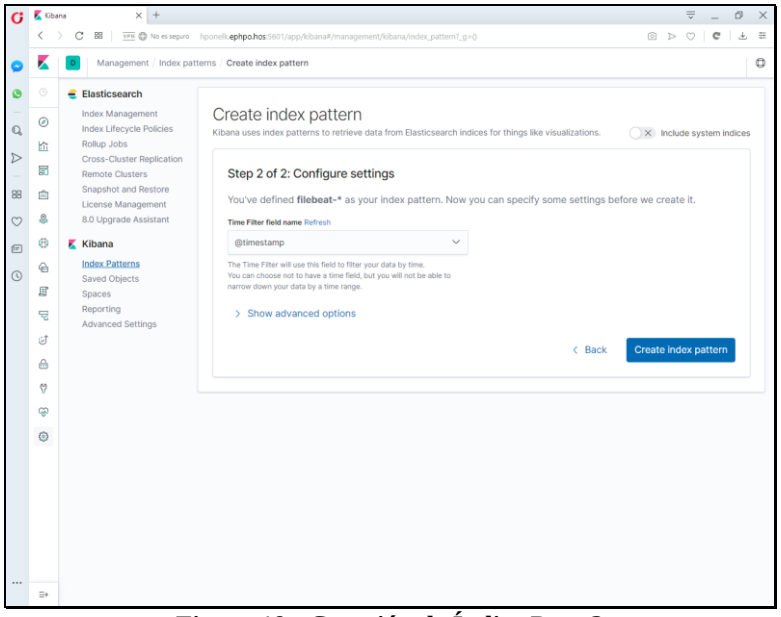


Figura 19 - Creación de Índice Paso 2

El filtro de tiempo elegido es "@timestamp" que es la fecha en la que se ha generado la entrada del log que proviene de Zeek (Bro IDS).

A continuación, vamos a comprobar que se están cargando datos desde ambos nodos, uno por cada centro instalado filtrando por el nombre de host del agente:

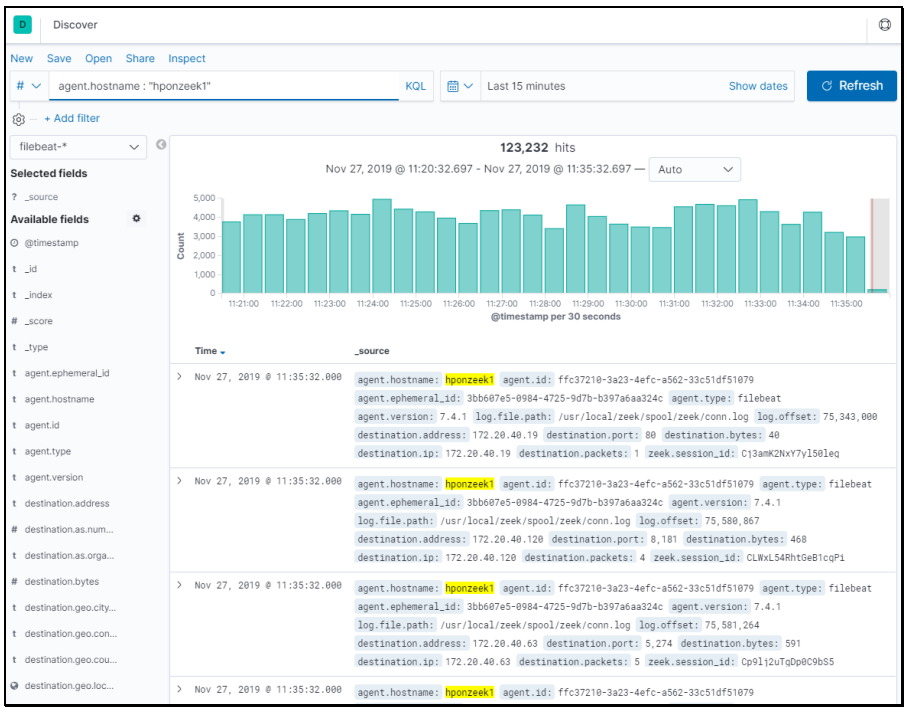


Figura 20 - Nodo Centro 1 "HPONZEEK1"

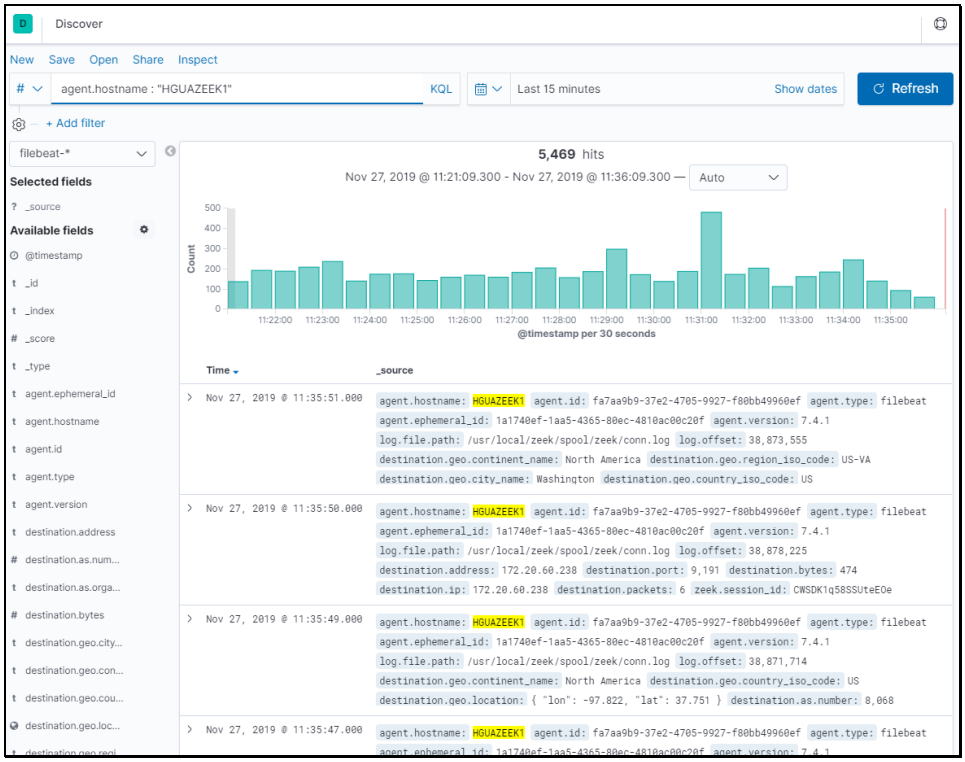


Figura 21 - Nodo Centro 2 "HGUAZEEK1"

En esta versión, 7.4.1, al integrar FileBeat con Kibana a través de la configuración de Filebeat, carga de forma automática unos objetos predefinidos que nos ayudarán a comenzar de forma más rápida el análisis del tráfico capturado aunque posteriormente tendremos que definir los propios nuestros dependiendo del tipo de anomalías que queramos detectar.

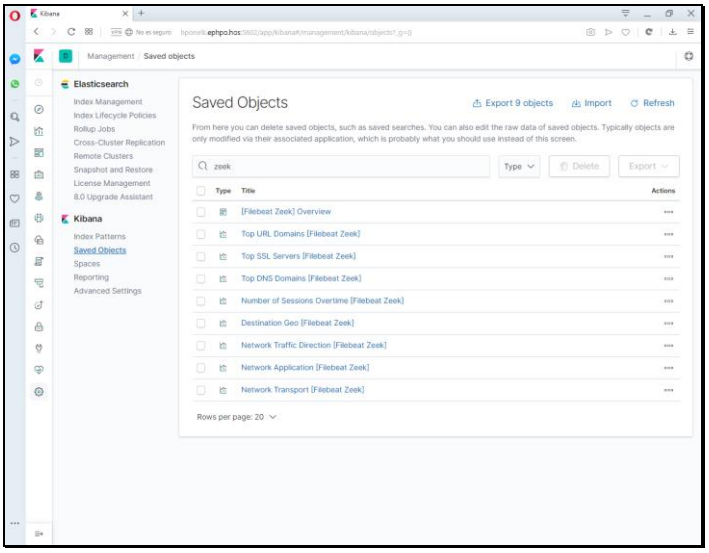


Figura 22 - Objetos creados por defecto para Zeek

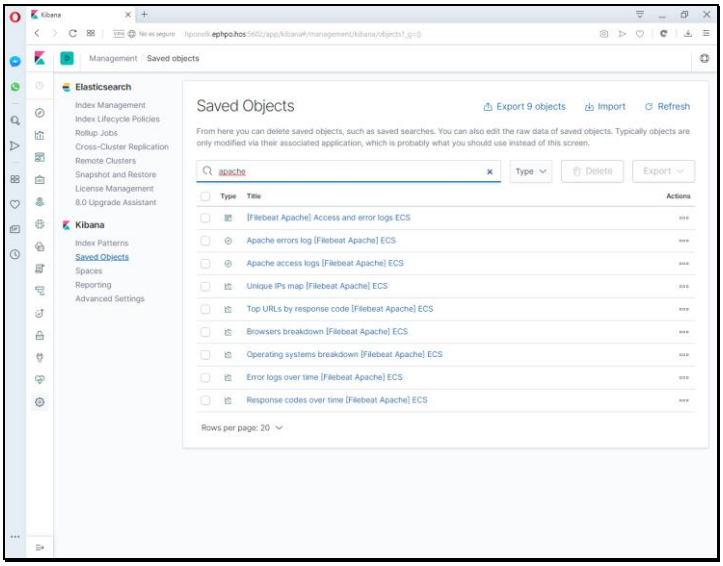


Figura 23 – Objetos creados por defecto para Apache

5.5. Integración SIEM

Como comentamos anteriormente, la ventaja de la utilización de ELK Stack es que ya dispone de un módulo específico de SIEM. Si utilizamos cualquiera de los Beats enumerados anteriormente, la integración se realiza automáticamente. En nuestro caso, como hemos utilizado Filebeat, no hemos tenido que realizar ningún esfuerzo adicional para la integración.

Para acceder al módulo, desde Kibana seleccionamos:

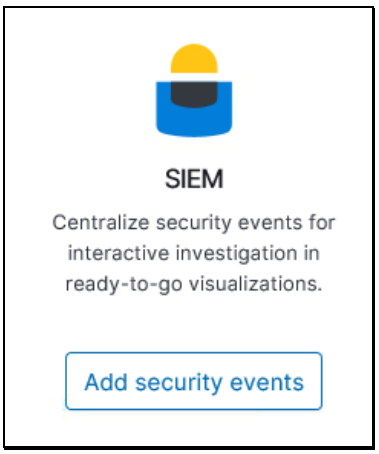


Figura 24 - Acceso módulo SIEM

A continuación, podemos observar que al utilizar Filebeat, concretamente el módulo Zeek de Filebeat, ya tenemos cargados una serie de eventos.

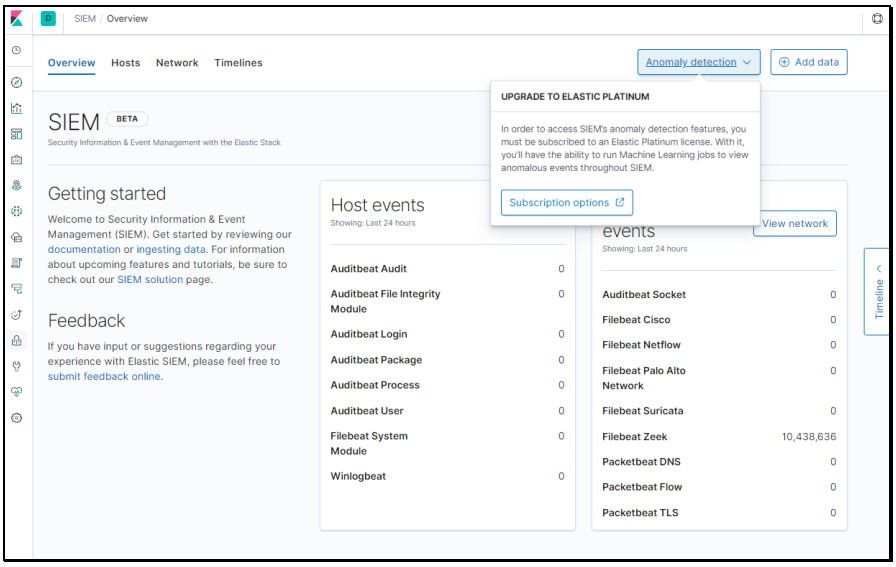


Figura 25 - Eventos Filebeat Zeek SIEM

También hay que destacar, que esta versión sólo dispone de la opción “Beta” y que el acceso a la “Detección Automática de Anomalías” está disponible en la licencia “Elastic Platinum” que nos permitiría la ejecución de trabajos de “Machine Learning” con realimentación al SIEM.

6. Producto final

Una vez realizada la instalación de los elementos principales que componen nuestro entorno de captura, transformación y visualización de datos, ahora debemos tener claro cuáles son, principalmente, los tipos de anomalías o comportamientos anómalos queremos tener especial atención teniendo en cuenta que el entorno en el que hemos realizado la implementación del sistema monitorizará el tráfico de red unos 350 equipos, en el caso de HGUAZEEK1, y unos 1600 equipos, en el caso de HPONZEEK1.

6.1. Test y Puesta en Marcha

Después de implementar cada uno de los componentes software a utilizar y definida la topología la topología de red, es necesario realizar una curva de aprendizaje para comprender cómo el producto final obtenido (la conjunción de todos los componentes instalados) puede ayudarnos a la detección de anomalías en la red.

Para ello, hemos dejado varias semanas funcionando todo el producto final para poder observar la correlación de eventos que se van originando. Hay que tener claro que el objeto de este trabajo no es encontrar todas las posibles amenazas o anomalías de la red sino comprender cómo nos puede ayudar a detectarlas y evidenciar que, el esfuerzo necesario para la implementación de los productos elegidos para este trabajo, nos ayudan a encontrar comportamientos anómalos que pueden indicar una amenaza de seguridad, siempre prestando mayor atención en aquellas que puedan poner en peligro la Seguridad del Paciente.

En definitiva, se podría intentar resolver inquietudes como:

- ¿Quién está atacando y dónde se ubica?
- ¿Cuál es la posible motivación de este ataque?
- ¿Qué puede pasar si el ataque tiene éxito?
- ¿Qué se pretende con este ataque?

Podríamos comenzar con algunos indicadores de red que son el común denominador frente a las posibles brechas de seguridad detectadas en un ambiente de producción como consumo anormal de ancho de banda, lentitud y tráfico inusual, etc. pero para comenzar, como ya hemos dicho que disponemos de datos de varias semanas capturados y almacenados, podemos iniciar nuestro primer análisis con los resultados que nos proporciona la herramienta de Machine Learnig que dispone ELK Stack.

Comenzamos con la ayuda del Machine Learnig

De esta forma, lo primero que vamos a hacer es acceder a la entrada de "Machine Learning" del menú de la izquierda y nos saldría una pantalla como la que sigue:

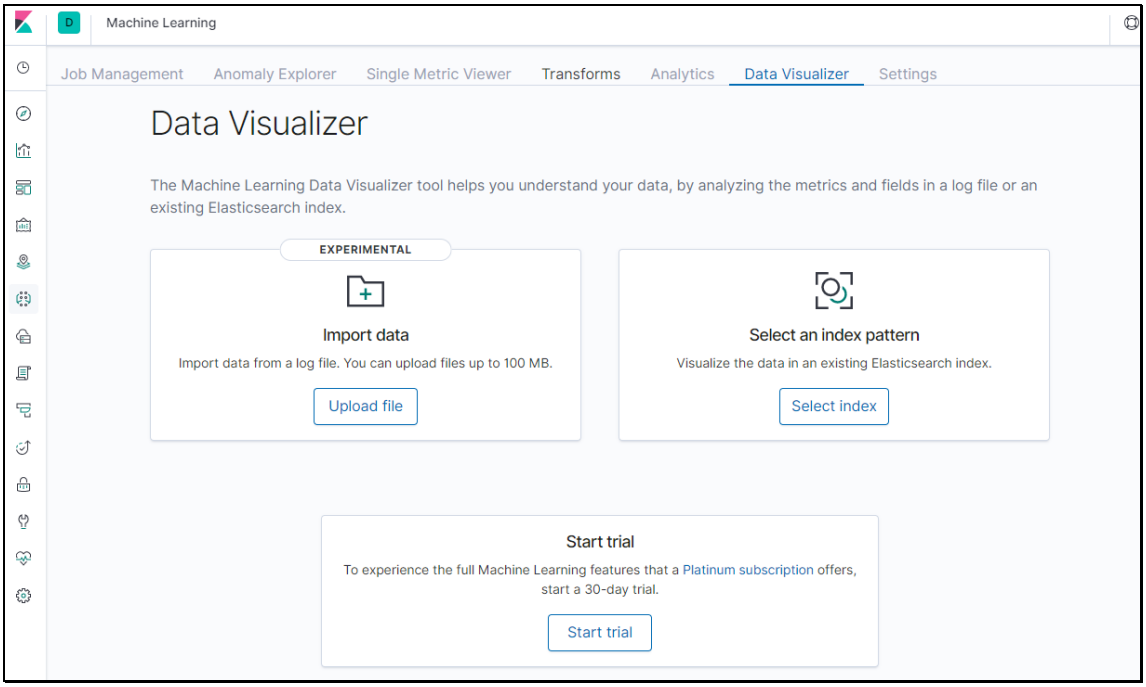


Figura 26 - Acceso Machine Learning ELK Stack

En la figura anterior, podemos observar que no todas las entradas están habilitadas y sólo es posible las de "Data Visualizer" y "Transforms" ya que la versión que utilizarmos es la gratuita (ver Anexo I) aunque será más que suficiente para ayudarnos con ciertas detecciones.

A continuación, seleccionamos el índice definido en anteriores capítulos (filebeat-*):

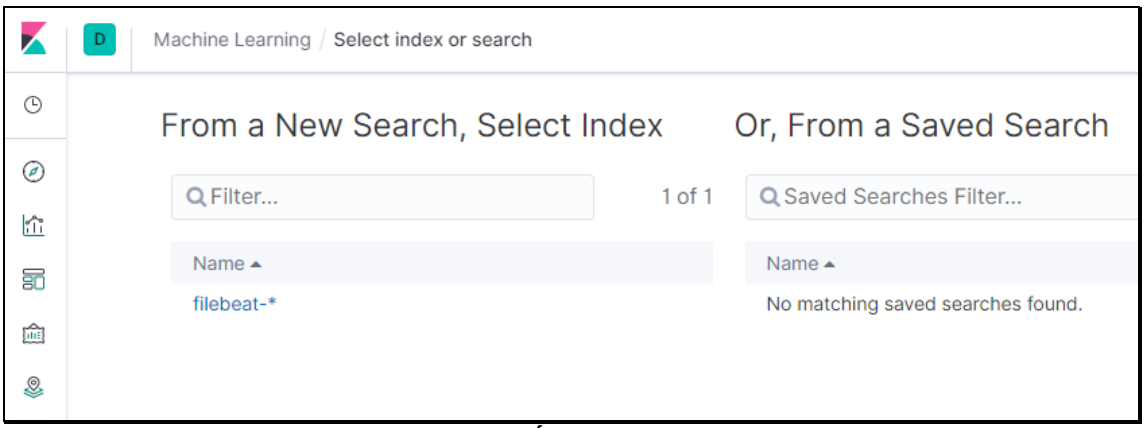


Figura 27 - Selección de Índice para Machine Learning

En el paso siguiente, podemos ver que nos sale una serie de visualizaciones donde, muy probablemente, sea necesario realizar un filtro debido a la gran cantidad de ellas que nos muestra.

De esta forma, para comenzar vamos a realizar un filtro para que solamente nos muestre aquellas visualizaciones que contengan la palabra “source” y posteriormente la palabra “destination”. Con ello pretendemos encontrar algunos comportamientos anómalos desde el origen (host o puerto) hacia varios destinos distintos y comportamientos anómalos desde cualquier origen a un destino (host o puerto) concreto, respectivamente.

Tras cargar todos los documentos disponibles, para una mayor cantidad de información agregada sobre la que sacar conclusiones lo más aproximadas posibles, obtenemos lo siguiente:

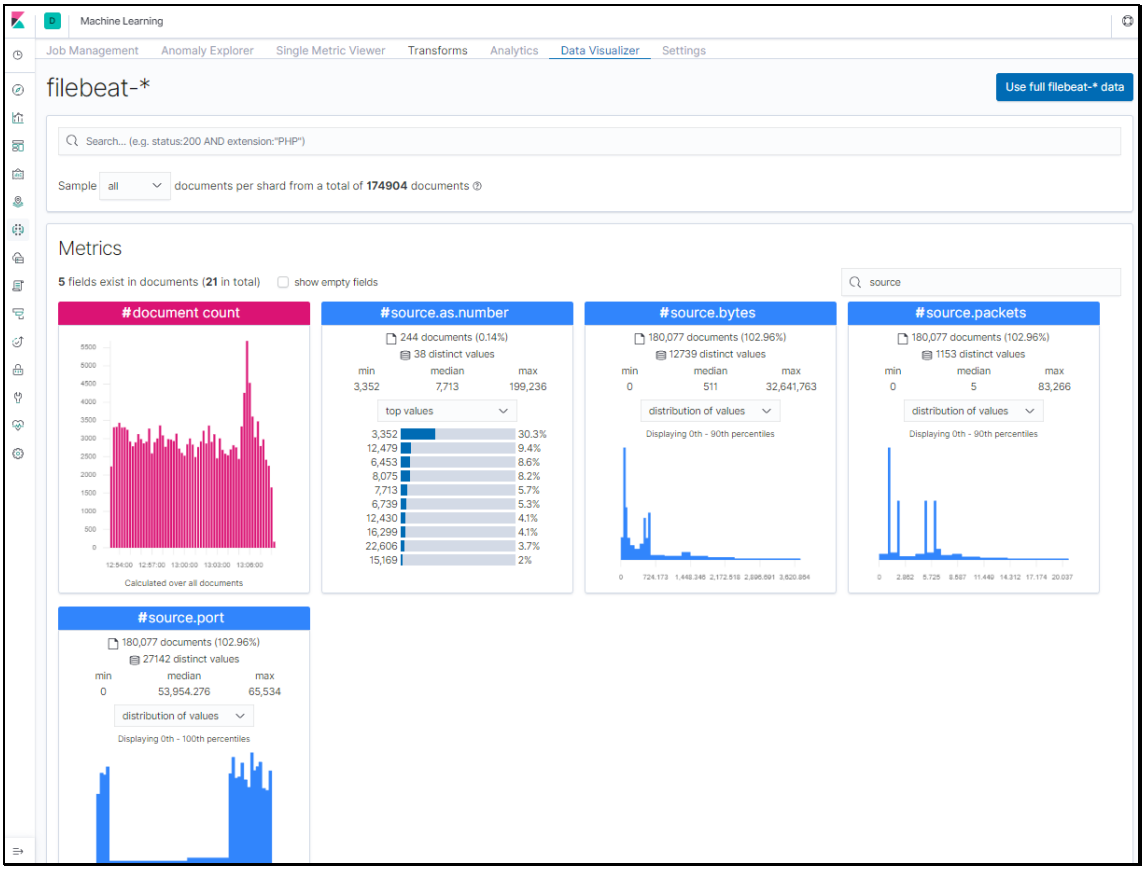


Figura 28 - Métricas "source" Machine Learning

Como primer análisis de este grupo de visualizaciones de tipo métrica, podemos determinar la gran mayoría del tráfico está compuesto por pequeños paquetes y cuyos puertos de origen está entre los 500 primeros (suelen ser puertos reservados del sistema operativo) y puertos a partir del 30.000 (puertos no reservados).

Continuando con la serie de gráficas, la siguiente nos muestra el análisis de los datos cargados por los campos cargados en el índice:

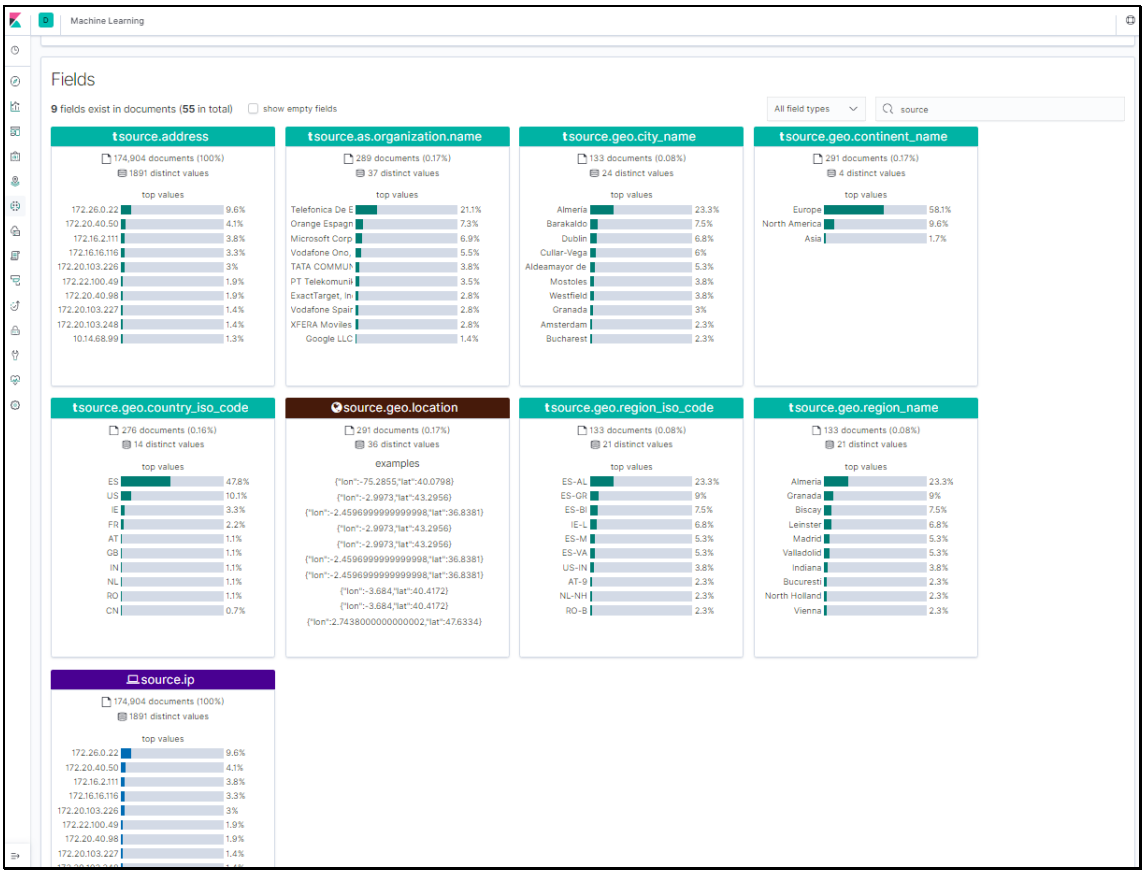


Figura 29 - Campos "source" Machine Learning

En esta figura anterior, podemos encontrar la primera de las anomalías que nos pueden indicar que algo raro está sucediendo. Concretamente:

- De más de 174.000 documentos analizados, un 9,4% de ellos contiene la IP de origen 172.26.0.22 que se corresponde con un equipo de Consultas Externas.

Después nos aparece la IP 172.20.40.50 que se corresponde con nuestro DNS Principal, con lo que es posible este tráfico tan importante.

Ahora continuamos con las visualizaciones y posterior análisis de las mismas para el filtro “destination”:

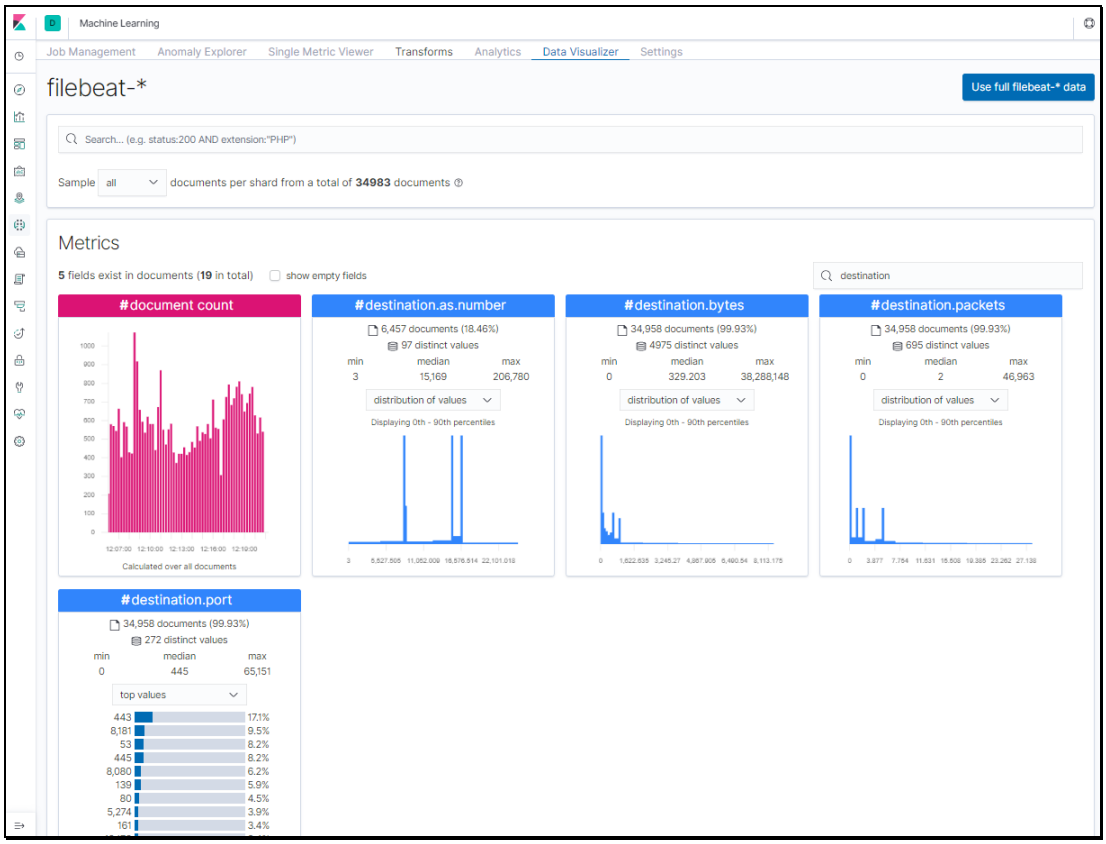


Figura 30 - Métricas "destination" Machine Learning

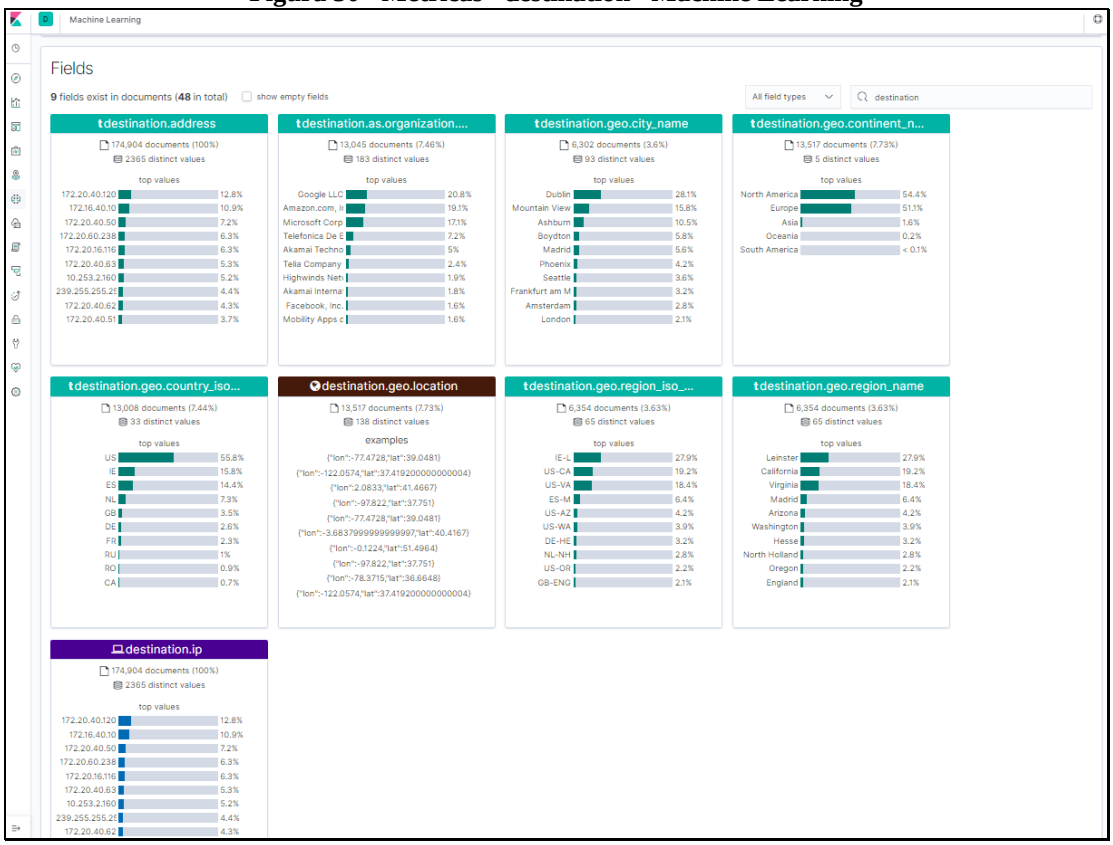


Figura 31 - Campos "destination" Machine Learning

Igualmente que para el caso anterior, una vez descartados los posibles comportamientos normales derivados de servidores (aquellos del rango 172.*.30.*, 172.*.40.* y proxies o enrutadores 172.*.16.*), se puede observar que los puertos con mayor tráfico de red o con más número de accesos, por este orden, son:

- 443 → Navegación SSL
- 8181 → Servidor MIRTH
- 53 → DNS
- 445 → SMB

Es extraño el número de accesos para SMB ya que, aunque se dispone de 1 servidor de ficheros por centro, las estadísticas de uso que se manejaban hasta el momento de la implantación de estas herramientas de detección no indicaban que tuviesen tanta carga de trabajo.

A continuación, se tomó la determinación de aislar el equipo que podía tener un comportamiento anómalo (172.26.0.22) y analizarlo detenidamente. Obteniendo los siguientes resultados:

- El equipo tenía instalado el antivirus corporativo pero, inexplicablemente, llevaba varios meses sin actualizar.
- Tras instalar una versión posterior del antivirus, se detecta que está infectado por Wannacry.
- Se comprueba que es un equipo con Windows XP SP3 y que dispone del parche de seguridad KB4012598 pero ha debido infectarse con anterioridad a la aplicación del mismo.
- Se toma la decisión de formatearlo y de emprender la consiguiente investigación por si ha infectado a otros equipos de la red.

Efectivamente, el tráfico que entendíamos anormal de la IP 172.26.0.22 y el correspondiente al puerto 445 (SMB) nos indicaba que existía una infección por Wannacry.

Investigación equipos infectados por Wannacry

Una vez llegados a este punto y con la infección confirmada, la siguiente fase nos lleva a intentar identificar si el equipo ha podido infectar a otros equipos de la red y qué consecuencias ha podido tener.

En el Antivirus Corporativo, se detectan varios equipos con intentos de infección:

Endpoint	Platform	Security Threat	Infected File/Object	Path of file	IP Address
GPOLI_85	Windows 7 6.1.7601	VBS_POWLOAD.GAC	MSShell32	G:\	172.16.0.85
HPONMAIL2	Windows Server 2012 R2 6.3.9600	TROJ_GEN.R032C0DDB19	md50000048571.msg (22010620_0120190409.exe)	E:\CORREO\ephpo.es\elenamaria.fernandez.guerrero\	172.20.40.98
FARMA3_235	Windows 7 6.1.7600	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.101.235
ALMR2_43	Windows 7 6.1.7600	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.2.43
PCEDIG9_58	Windows 7 6.1.7601	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.104.58
PGES230	Windows 7 6.1.7600	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.0.230
PTRAUS9_75	Windows 7 6.1.7600	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.104.75
PSMI103	Windows 7 6.1.7600	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.0.103
FARMA4_84	Windows 7 6.1.7600	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.102.84
PREGISTRO49	Windows 7 6.1.7601	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.0.49
PTRAUCURAS9_29	Windows 7 6.1.7601	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.104.29
PMPREV2_99	Windows 7 6.1.7600	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.2.99
PENDOS_209	Windows 7 6.1.7601	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.1.209
FARMA3_236	Windows 7 6.1.7600	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.101.236
PCARDIO9_31	Windows 7 6.1.7601	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.104.31
FARMA4_83	Windows 7 6.1.7600	WORM_WCRY.D	mssecsvc.exe	C:\WINDOWS\	172.20.102.83

Figura 32 - Log Antivirus Corporativo

Para poder detectar un tráfico de red anómalo utilizando ELK Stack tenemos varias alternativas. Como ya hemos indicado, nos puede ayudar Machine Learning y también algunas visualizaciones predefinidas que ya mencionamos en capítulos anteriores. Para el caso que nos ocupa, posibles equipos infectados por Wannacry, tenemos que detectar el comportamiento anómalo que ocasiona esta infección.

Según el Incibe (Instituto Nacionad de CiberSeguridad) se recomienda bloquear la comunicación de los puertos 137/UDP y 138/UDP así como los puertos 139/TCP y 445/TCP en las redes de las organizaciones para evitar la propagación; esto es, porque son las vías principales de propagación. Por ello, se decide crear una visualización para la detección concreta de comportamientos anómalos en el puerto 445 con las IPs (Top Ten) que mayor número de entradas tengan como destino el puerto indicado.

Consultando la web oficial de Avast (conocido antivirus), el ransomware WannaCry ataca a las redes usando SMBv1, un protocolo que ayuda a los equipos a comunicarse con las impresoras y otros dispositivos conectados a la red. Esta versión, diseñada en 2003, dejó a los equipos expuestos a los hackers y esta vulnerabilidad se denominó MS17-010.

Tomando como referencia el índice, en el eje Y aparecerá el número de entradas o documentos al puerto 445 y en el eje X la IP a la que corresponde el agregado de entradas definidas. Además se hace en orden descendente al número de entradas por IP con un límite de aparición de 10 IPs distintas, quedando de la siguiente forma:

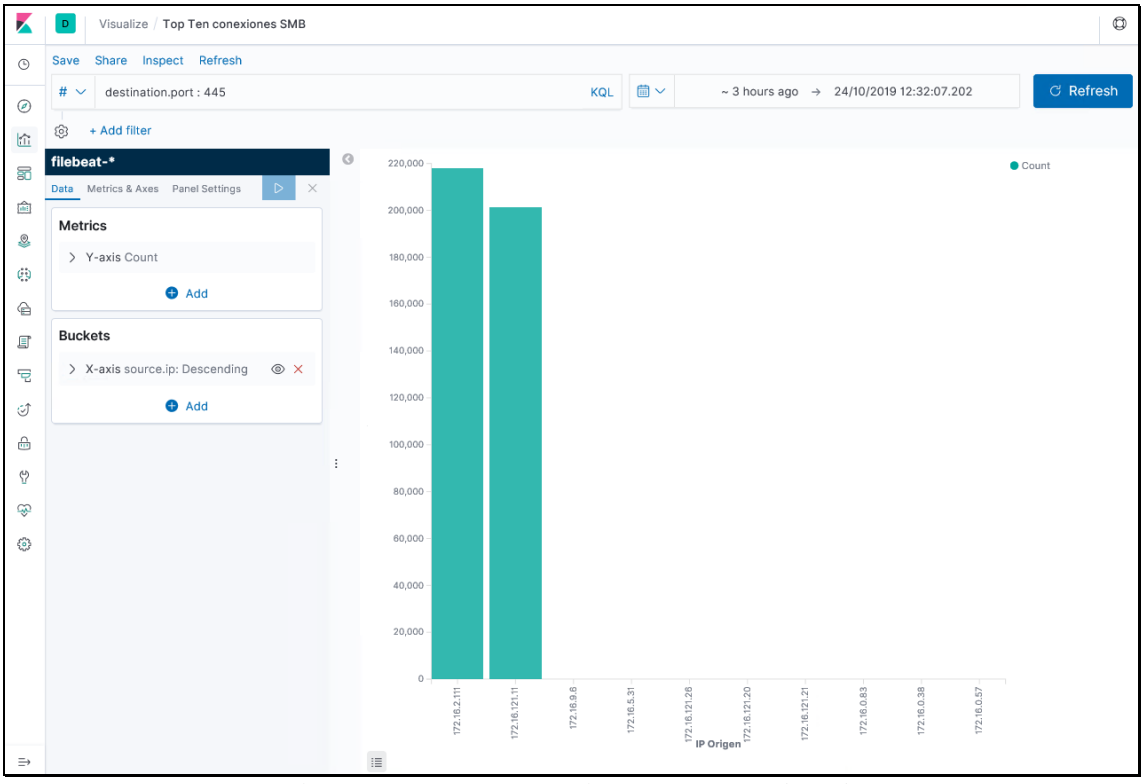


Figura 33 - Top Ten conexiones puerto 445 (SMB)

Si observamos la gráfica anterior, filtrado por las últimas 3 horas, vemos que hay 2 IPs que destacan de forma considerable, con más de 200.000 entradas, respecto al resto. Concretamente las IPs 172.16.2.111 y 172.16.121.11, pertenecientes a rango de IPs de equipos electromédicos y equipos de laboratorio, respectivamente.

Precisamente estos equipos son objetivo prioritario de este trabajo ya que podrían suponer un riesgo para la Seguridad del Paciente debido a la gestión de información que pueden manejar o la repercusión que pueden tener debido a un mal funcionamiento. Por ello, para se realizará una investigación más detallada de la anomalía para cada equipo.

Equipo 172.16.2.111

Se determina que es un equipo electromédico que gestiona el stock y la imputación de lentes oculares a los pacientes que son operados de "cataratas oculares" con lo que de forma inmediata se procede a su desconexión de la red, quedando aislado, y se notifica la sospecha de infección a la empresa que lleva el mantenimiento y soporte.

Tras la notificación a la empresa externa, se pone en contacto un técnico especialista confirmando que el equipo tiene embebido un Windows 7 y que, generalmente, no se suelen aplicar de forma periódica los correspondientes parches de seguridad y que tampoco dispone de antivirus instalado y que procederán a realizar visita para su actualización y verificación de la infección.

Por otro lado, aprovechando otra visualización disponible en Kibana, ya comentada anteriormente, procedemos a observar a qué direcciones IPs se intenta conectar para infectar resultando lo siguiente:

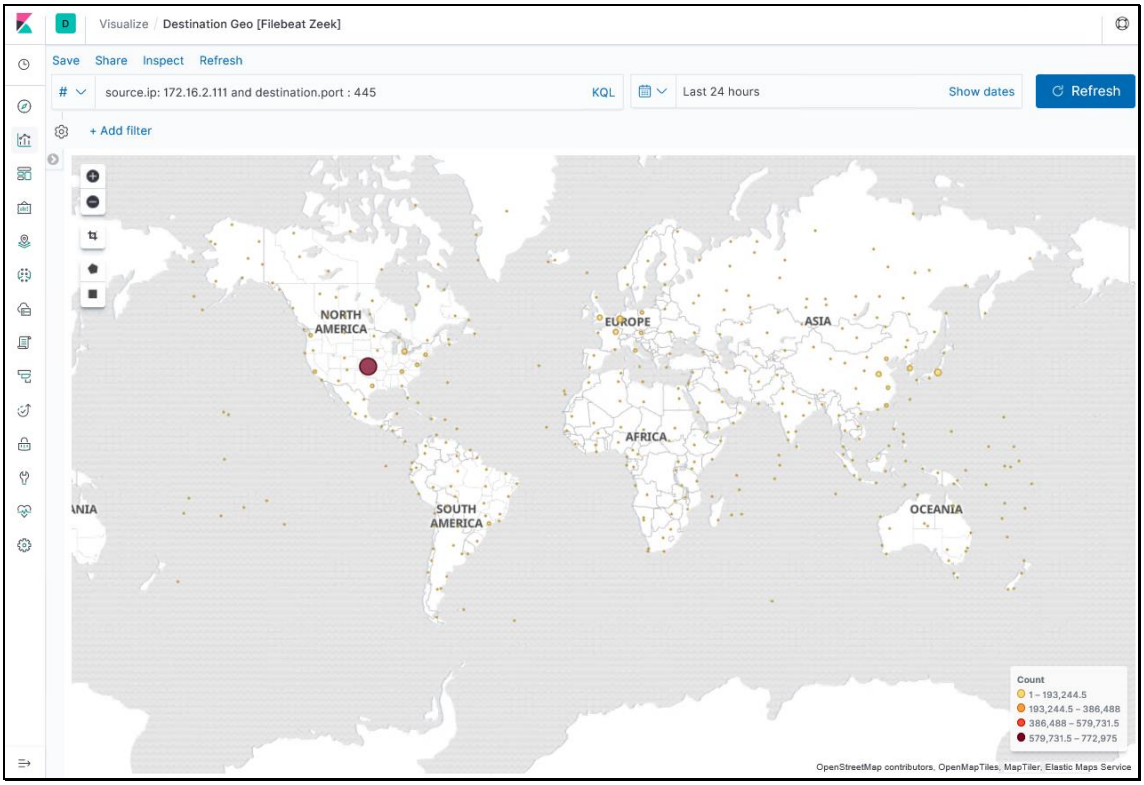


Figura 34 - Mapa Mundial intentos infección Wannacry

Podemos observar que el equipo que estamos investigando, en las 24 horas anteriores, ha intentado conectar a direcciones IPs de todo el mundo al puerto 445, especialmente de Norte América.

Tras la visita del técnico, se confirma, con la instalación de un antivirus, que el equipo está infectado con Wannacry y se procede a actualizarlo y limpiarlo.

Una vez confirmado este hecho, cobra especial relevancia la preocupación por si esta infección ha podido, de alguna forma, ocasionar un funcionamiento anómalo del equipo que pueda afectar a la Seguridad del Paciente.

¿Qué comportamiento tiene el gusano Wannacry?

Para ello, volvemos a consultar la página oficial de Avast y nos indica que este tipo de ransomware se comporta como un gusano: se propaga por las redes hasta llegar a su equipo y, una vez allí, cifra sus archivos. Al infectarlo, recibe la advertencia de que no podrá acceder a sus archivos o, peor aún, no podrá ni abrir la sesión en el equipo.

¿Qué funciones realiza el equipo electromédico infectado?

Consultado la documentación aportada por la empresa que instaló el equipo, éste dispone de un software específico que realiza una consulta a la base de datos con los datos del paciente al que se le va a implantar la lente ocular y, siguiendo las instrucciones del oftalmólogo, se hace una imputación de forma manual qué lente se le ha implantado, dándola de baja del inventario. Es un proceso manual que no guarda ningún fichero local y que no ha podido verse afectado por el tipo de infección detectada.

Equipo 172.16.121.11

Se determina que esta dirección IP se corresponde con un equipo de análisis de muestras de laboratorio y se procede a notificar el comportamiento anómalo y sospechas de que ha podido ser infectado por el conocido gusano a la empresa encargada de su mantenimiento y se procede a aislarlo.

Esa misma mañana, un técnico se desplaza a las instalaciones y confirma que el equipo dispone de Windows 7 instalado desconociendo desde cuándo no se han instalado parches de seguridad. Al instalarle un antivirus se confirma las sospechas de infección. Se procede a la instalación de todas las actualizaciones de seguridad disponibles y al escaneo completo con antivirus confirmando que en el equipo no se muestran nuevas alertas de infección.

De la misma forma que en el caso anterior, procedemos a investigar si esta circunstancia ha podido conllevar un riesgo para la Seguridad del Paciente.

¿Qué funciones realizar el equipo infectado?

El equipo procesa las muestras recibidas en tubos etiquetados por códigos de barras y devuelve los resultados obtenidos mediante protocolo estándar de comunicación con la base de datos de Laboratorio insertando en las tablas correspondientes los valores asociados al paciente que se le ha extraído la muestra.

Tras consultar con responsable del departamento de Laboratorio, la empresa encargada de la instalación, mantenimiento y funcionamiento del equipo electromédico y conociendo el comportamiento del gusano, que sólo encripta ficheros y no afecta a procedimientos relacionados con conexiones a bases de datos, se determina que la infección no ha ocasionado un incremento del riesgo en la Seguridad del Paciente más allá de la posible parada de funcionamiento que, en determinados casos, puede ocasionar la infección por este ramsomware.

Ramsomware Bluekeep

Coincidiendo con la realización de este trabajo, se produce una oleada de ataques para aprovechar la vulnerabilidad apodada Bluekeep (CVE-2019-0708) y Microsoft realiza una campaña masiva de información acerca de la importancia de parchear ante la factible posibilidad de que en el futuro cibercriminales hagan uso de esta vulnerabilidad para distribuir amenazas más peligrosas.

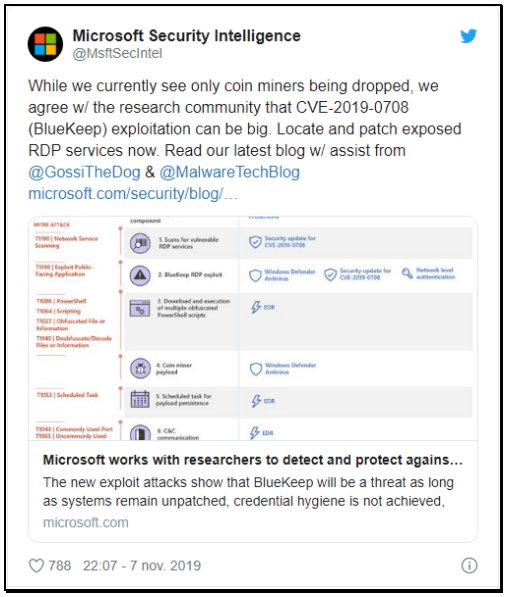


Figura 35 – Campaña en Twitter de Microsoft sobre Bluekeep

Para obtener más información nos dirigimos, como hicimos anteriormente, a la página web oficial del antivirus Avast obteniendo la siguiente información:

"¿Qué es BlueKeep?

BlueKeep es una vulnerabilidad de software que afecta a versiones antiguas de Microsoft Windows. El riesgo es significativo porque ataca el Protocolo de escritorio remoto (RDP) del sistema operativo, que permite conectar con otro ordenador a través de una conexión de red. De este modo, una ciberamenaza podría extenderse a gran velocidad. Se detectó por primera vez en mayo, en el Centro Nacional de Ciberseguridad del Reino Unido. Desde mediados de mayo, Microsoft viene suplicando con mensajes muy poco sutiles a los usuarios amenazados (alrededor de un millón) que apliquen el parche correspondiente.

¿Me afecta a mí?

Podría ser, si hace tiempo que no actualizas el software de tu ordenador personal. Microsoft dice que los sistemas vulnerables con soporte (los que todavía reciben soporte técnico de la empresa) son Windows 7, Windows Server 2008 R2 y Windows Server 2008. Entre los sistemas sin soporte están Windows 2003 y Windows XP. Los clientes con Windows 8 y Windows 10 no se ven afectados por esta vulnerabilidad.

¿Qué debo hacer al respecto?

Debes descargar y aplicar el parche, o actualización de software, que resuelve la vulnerabilidad. Puedes encontrar las descargas para las versiones de Windows con soporte en la Guía de actualización de seguridad de Microsoft. Los clientes que utilicen versiones con soporte de Windows y que tengan habilitadas las actualizaciones automáticas estarán protegidos automáticamente.”

De esta forma, se procede a la liberación del parche de seguridad en todos los equipos y servidores de la empresa y la notificación a las empresas externas encargadas del mantenimiento de equipos electromédicos para que, de forma manual, apliquen la actualización correspondiente.

No conforme con ello, aprovechando la implantación de nuestros productos, definimos una nueva visualización, “Top Ten conexiones RDP”, para detectar posibles comportamientos anómalos específicamente para el puerto 3389 (RDP), al igual que hicimos para el puerto 445 (SMB).

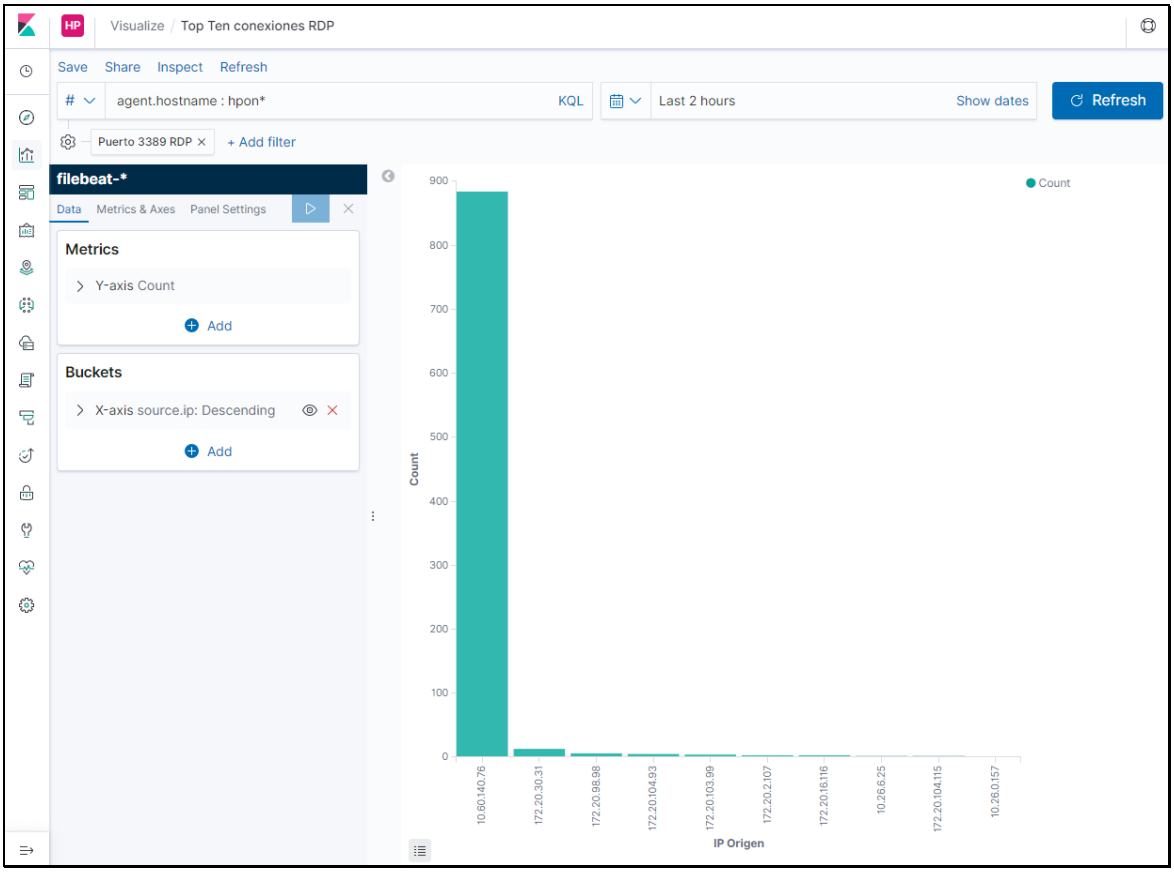


Figura 36 - Posible infección por Bluekeep

Como podemos observar en la figura anterior, hay una determinada dirección IP externa a la empresa, 10.60.140.76, que está realizando un número de conexiones muy elevado al puerto de conexión de escritorio remoto, concretamente unas 900 conexiones en 2 horas. Vemos que es un comportamiento anómalo si lo comparamos con el número de conexiones, en el mismo intervalo de tiempo, que realizan las otras 9 direcciones IPs del “Top Ten”.

De esta forma, comenzamos de nuevo otra nueva investigación para determinar a qué equipo se corresponde esta IP y si, efectivamente, se corresponde, como en los casos anteriores, con una infección por ramsonware, en este caso por Bluekeep.

Esta investigación no lleva a que la IP se corresponde con la conexión de un equipo de fuera de nuestras instalaciones perteneciente a una empresa externa en la que se ha contratado la realización de pruebas diagnósticas de Resonancia Magnética Nuclear. Evidentemente, para la conexión a nuestras instalaciones se han aplicado reglas de nuestro firewall que permiten este tipo de conexiones y se procede a notificar la posible infección, cortando de forma inmediata el tráfico en nuestro firewall.

Minutos más tarde, se pone en contacto un técnico informático de la empresa de Resonancia Magnética indicando que no se trata de una infección sino de una monitorización que han establecido para comprobar que el puerto 3389 se encuentra activo y, en caso contrario, generar una alerta que les permita proceder a contactar con nuestro departamento informático para subsanar la incidencia. Una vez detenida esta monitorización y activada de nuevo la regla del firewall se comprueba que ya no se detecta este comportamiento anómalo.

Se determina que, aunque se ha detectado un comportamiento anómalo, éste no ha sido derivada de ninguna infección pero sí que nos puede ayudar a detectar, en un futuro, comportamientos similares que sí lo sean.

Caída de servidor Oracle

Durante varios días consecutivos se están produciendo paradas inesperadas en nuestro servidor principal de Base de Datos que está ocasionando cierto malestar entre los profesionales sanitarios y, evidentemente, en la dirección de los distintos hospitales.

Con las herramientas de gestión con licencia que se dispone actualmente en la empresa y habiendo gestionado la correspondiente incidencia con la empresa externa que da soporte a nuestro Sistema Gestor de Base de Datos Oracle, no se ha podido determinar la causa o causas que están ocasionado estas paradas.

Los servidores en los que se aloja la base de datos se encuentran en proceso de licitación para su renovación debido a que ha llegado su “End of Support”, determinado así mediante la publicación del boletín semestral que hace IBM respecto al equipamiento que quedará sin soporte a partir de una fecha, recomendando su renovación.

Así, aprovechando la implementación del producto sobre el que versa este trabajo y cuya implementación ya ha sido finalizada, se decide crear una nueva visualización, igual que en los casos descritos anteriormente, llamada “Top Ten Conexiones Oracle” para intentar detectar si ha podido haber alguno o varios comportamientos anómalos que hayan podido contribuir a la caída y reinicio de nuestro servidor mediante la conexión al puerto 1521, que es el puerto estándar de conexión a Oracle.

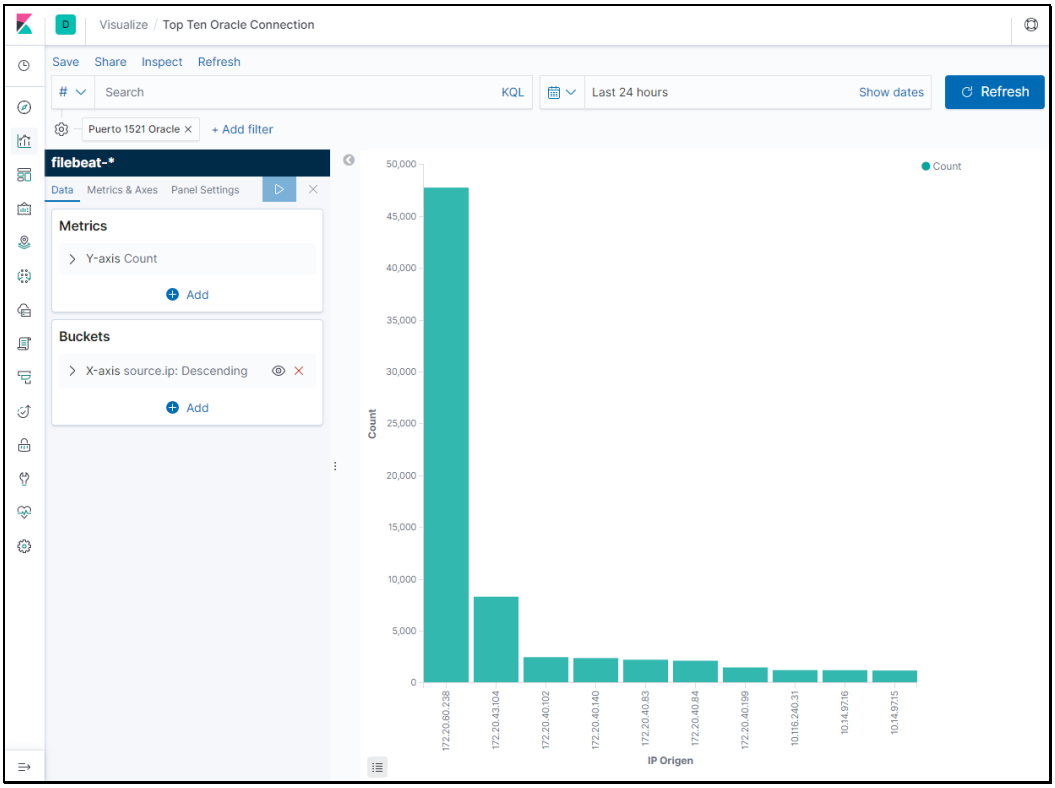


Figura 37 - Conexiones excesivas a Oracle ocasionando reinicio del servidor

De esta forma, tras establecer como período de visualización las últimas 24 horas, podemos observar que desde una dirección IP, 172.20.60.238, se están produciendo cerca de 50.000 conexiones, muy superior a la segunda dirección de red que no alcanza las 10.000 conexiones.

Se determina que el equipo se corresponde con el servidor de impresión instalado por una empresa externa y cuyo software, PaperCut, se conecta a nuestra base de datos para llevar la contabilidad de número de copias para la facturación mensual establecida mediante contrato de renting de equipos de impresión y multifunción.

Tras contactar con los técnicos de la empresa auxiliar, nos informan que, en días pasados, procedieron a realizar una actualización del software de gestión de contabilidad de copias, PaperCut, y se concluye que esta actualización puede estar ocasionando un crecimiento excesivo en el número de conexiones concurrentes que pueda estar provocando el consumo excesivo de recursos en nuestro servidor de base de datos, algo antiguo, y su reinicio imprevisto.

Con estos datos, se procede a la reversión de la actualización del software indicado y durante los días posteriores no se detecta ningún comportamiento anómalo de conexiones al puerto 1521 y tampoco se producen reinicios inesperados.

Riesgo para el paciente

Analizando este caso particular, la caída del sistema gestor de base de datos ORACLE, ha ocasionado que, durante algunos minutos, el acceso a la Historia Digital del paciente en la que se incluye, entre otros, el Historial Médico, información sobre alergias, analíticas y prescripciones de medicamentos no ha sido posible y ha podido interferir en la atención médico-hospitalaria.

Esta incidencia ha podido ocasionar un riesgo para algunos pacientes a los que no ha sido posible aplicar el tratamiento oportuno o, como caso más remoto, haya podido derivar en un incorrecto diagnóstico con lo que se notifica el incidente a los responsables sanitarios para que se tomen las medidas oportunas.

6.2. Otras configuraciones

En este caso, nos centraremos en la ayuda que nos proporciona Kibana para poder organizarnos mejor y ayudarnos a obtener una visión más clara acerca de las anomalías que se puedan estar produciendo en nuestra red y poder detectarlas.

Definición de espacios de trabajo

La organización es primordial para muchos usuarios de Kibana, en especial para aquellos con cientos de dashboards y visualizaciones. Es una característica disponible desde la versión 6.5 de Kibana y permite agrupar y organizar los dashboards, visualizaciones y otros objetos guardados en compartimientos separados.

En este trabajo se han definido 2 espacios de trabajo que se corresponden con los distintos centros donde se han instalado las sondas con Zeek, así hemos conseguido aislar el contenido en áreas separadas para:

- Disminuir el desorden.
- Facilitar el descubrimiento.
- Evitar duplicados de información por accesos entre ambos centros a través de la Red Virtual (VPN) de interconexión.

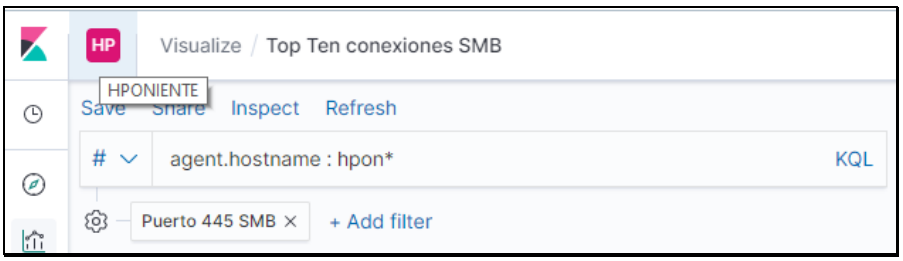


Figura 38 - Ejemplo Espacio HPONIENTE

Definición de Dashboards

Para obtener una visión más rápida de lo que está sucediendo, podemos agrupar cada una de las visualizaciones que hemos definido o las que, como hemos indicado, ya vienen por defecto. Un panel de Kibana es una colección de visualizaciones, búsquedas y mapas, generalmente en tiempo real.

Los paneles proporcionan información de un vistazo de sus datos y le permiten profundizar en los detalles.

Como ejemplo, podemos ver a continuación uno de los dashboard definidos en los que hemos incluído las visualizaciones anteriores que hemos utilizado para la detección de las anomalías y que, en algunos casos, derivaban de infecciones ocasionadas por Ramsomware, junto con otras visualizaciones por defecto que ya teníamos disponibles:

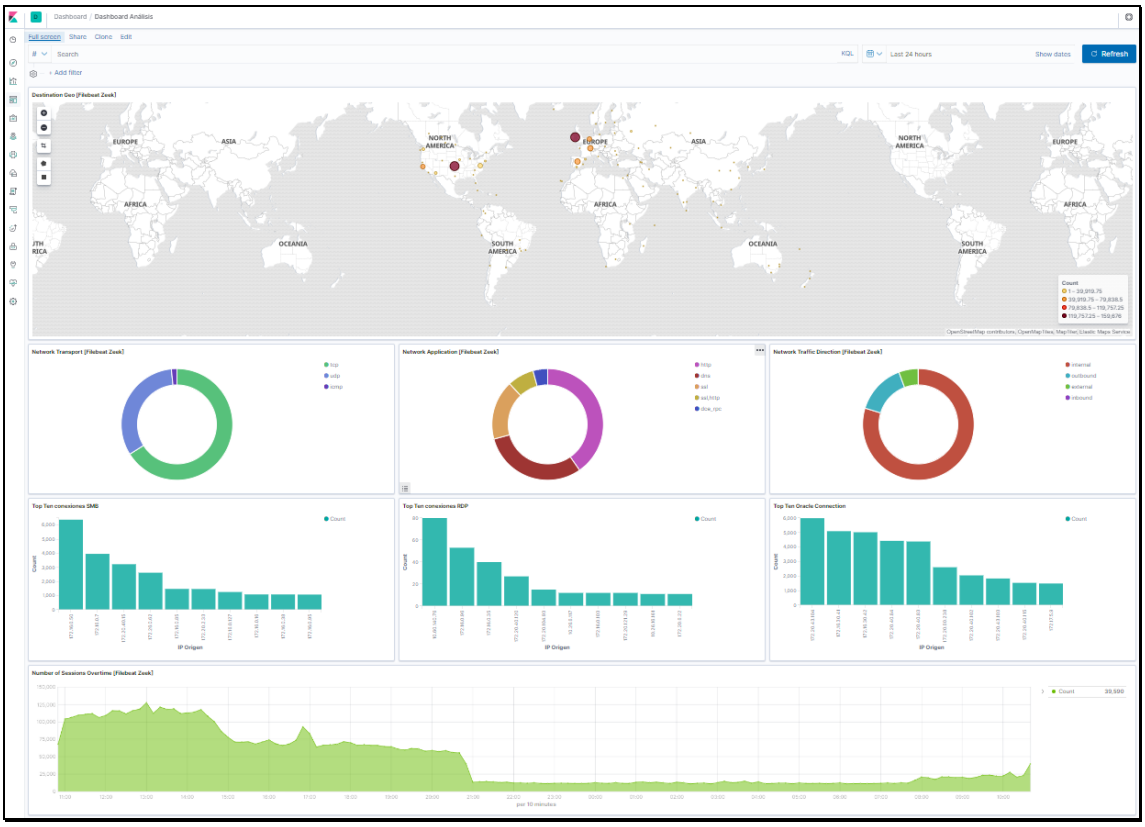


Figura 39 – Ejemplo Dashboard definido

De un solo vistazo podemos observar ciertos comportamientos comparando de forma directa tanto en tiempos definidos como en tiempo real, estableciendo filtros por cualquier campo de nuestro índice, por localización, por servicio de red, etc.

Trabajos de Rollups

Tan importante es la monitorización y obtención de datos como las tareas de mantenimiento asociadas a que nuestro sistema de monitorización siga activo. Por ello, se ha definido “trabajos de rollup o resúmenes”.

Estos trabajos son tareas periódicas que agregan datos de índices especificados por un patrón de índice y los acumula en un nuevo índice. Los índices acumulativos son una buena forma de almacenar de forma compacta meses o años de datos históricos para su uso en visualizaciones e informes. Ha sido necesario ya que estábamos recogiendo un volumen de información semanal superior a los 20GB.

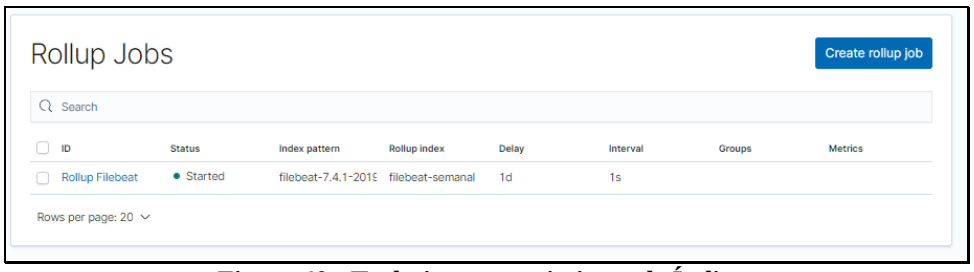


Figura 40 - Trabajos mantenimiento de Índices

7. Conclusiones

7.1. Consecución de objetivos marcados

Como primer paso para redactar las conclusiones, es necesario comenzar con la revisión de los objetivos del trabajo que inicialmente nos propusimos.

- ✓ **Identificar y describir cómo puede afectar una intrusión en un equipo electromédico en la seguridad del paciente.** → Se ha definido qué es la seguridad del paciente y algunas de las consecuencias que puede tener un comportamiento anómalo en un equipo electromédico, incluida una intrusión.

- ✓ **Identificar y describir cómo puede afectar una intrusión en la base de datos de historia clínica en la seguridad del paciente.** → Se ha identificado como uno de los principales errores que suelen producirse en relación a la seguridad del paciente los producidos en la prescripción de la medicación. En este caso, una modificación malintencionada ocasionada por un acceso no autorizado a la base de datos que pueda cambiar un parámetro en las dosis o incluso el identificador del principio activo a utilizar podría ocasionar graves problemas en la salud.

- ✓ **Analizar y comprender cómo puede ayudar un sistema de detección de actividades sospechosas en la identificación de las amenazas para el problema que se nos plantea en este trabajo.** → Se han detectado distintas actividades anómalas que han derivado a la detección de distintas infecciones de las que hemos determinado que han sido un potencial peligro para la seguridad del paciente ya que podrían haber ocasionado paradas inesperadas en nuestros sistemas.

- ✓ **Analizar y comprender cómo el Machine Learning puede ayudar en la detección de actividades sospechosas y su aplicación en nuestro sistema a implementar.** → Hemos definido qué es el Machine Learning y cómo nos puede ayudar a detectar actividades sospechosas que se salen de los comportamientos que podríamos considerar como "normales" si bien en cierto que debido a las restricciones de la licencia gratuita que hemos utilizado, no hemos podido aprovechar todo el potencial que nos ofrece.

- ✓ **Estudiar y aplicar sistemas de almacenamiento de gran volumen de datos para la ayuda en la identificación de actividades sospechosas en la red.** → Hemos utilizado Elasticsearch, basado en documentos JSON, como base de datos para el manejo de gran volumen de información y su posterior explotación en la detección de anomalías en nuestra red.

- ✓ **Estudiar y decidir qué sistema de detección de intrusiones es el adecuado a la problemática planteada y qué herramientas adicionales son necesarias para obtener el resultado deseado.** → Hemos documentado, analizado y decidido sobre la mejor alternativa que, en combinación, mejor se adaptaba a las características que necesitábamos para la realización de nuestro trabajo con la consecución de los objetivos marcados.

- ✓ **Investigar distintas fuentes externas de listas de reputación que puedan ayudar en la detección de actividades sospechosas y su aplicación en la implementación de nuestro sistema.** → Se han analizado los distintos riesgos y establecidos numerosos filtros basados en comportamientos anómalos referidos a infecciones de ramsomware conocidas y sobre las que, en el momento de realización del trabajo, hay campañas activas de infección. Es cierto que se ha conseguido este apartado solamente de forma parcial ya que no se ha establecido esta relación de forma automatizada.

- ✓ **Exponer las conclusiones del trabajo realizado.** → En los próximos párrafos.

7.2. Conclusiones del trabajo

Cuando nos planteamos un proyecto de investigación, nos marcamos una serie de objetivos muy ambiciosos ya que, en la mayoría de las ocasiones, partimos del desconocimiento de hasta dónde podemos llegar.

Conforme se ha ido avanzando en el estudio y la implementación he podido aprender que las herramientas elegidas nos pueden brindar infinitas posibilidades de monitorización y, en consecuencia, de detección de comportamientos que difieren del que tienen otros dispositivos similares. Así se puede concluir que son comportamientos anómalos o anomalías y que pueden ser indicio de alguna infección o amenaza.

Se ha demostrado que es posible:

- Implementar un ambiente de monitorización red mediante el uso de herramientas de software libre o gratuito que le permiten a una pequeña o mediana empresa obtener una visualización de lo que está ocurriendo dentro de su infraestructura, tanto de forma inmediata como analizando sucesos pasados, con el fin de contrarrestar intentos de intrusión, ataques en curso, infecciones y otras anomalías ocasionadas por un mal funcionamiento de tareas o del software.

- Obtener un servicio de valor agregado basado en la monitorización del tráfico de red, para que muchas empresas puedan proteger su información y sus activos tanto de forma preventiva como reactiva.

- Dar visibilidad a la importancia que tiene aplicar controles de seguridad tanto al tráfico entrante y saliente como al tráfico que circula en nuestra red, ya que nos puede dar mucha información sobre comportamientos que pueden ocasionar un futuro mal funcionamiento y así reaccionar oportunamente en una fase temprana frente a una situación anómala.

- Reutilizar recursos hardware que nos permiten afrontar un proyecto de seguridad a coste muy reducido o con valor residual cero.

- Reorientar las anomalías detectadas para intentar identificar riesgos concretos que pueden ser vitales tanto para la organización como para los usuarios que, de alguna u otra forma, intervienen en los procesos de cualquier empresa. En nuestro caso, orientados a la detección de anomalías que pudiesen afectar a la Seguridad del Paciente.

Este proyecto puede utilizarse como base para el desarrollo más extendido de un centro de monitorización empresarial. Para ello, se necesitará explicar la viabilidad y efectividad del proyecto a la alta dirección para poder arrancar el compromiso para facilitar los recursos tanto para poder afrontar la adquisición de licencias de uso empresarial como para la dedicación de resursos humanos y tecnológicos necesarios para realizar la implantación más avanzada para la detección de anomalías.

La valoración global del trabajo es muy positiva ya que hemos podido conseguir prácticamente todos los objetivos que inicialmente nos marcamos y la experiencia vivida ha sido satisfactoria aunque, debido a la dificultad que confiere la curva de aprendizaje de Elastic Stack, algo dura. Realizar una planificación justo al comienzo y analizar los principales riesgos que me podría encontrar durante el desarrollo del trabajo me ha ayudado a comprenden la dificultad que podría entrañar conseguir el hito más importante que, en este caso, es la entrega a tiempo de toda la memoria final.

Siendo conscientes que queda todavía mucho trabajo por hacer hemos establecido una metodología y unas herramientas que nos ayudarán a seguir avanzando en la detección de anomalías en nuestra red.

7.3. Líneas de trabajo futuro pendientes

Debido a las limitaciones en recursos disponibles y, sobre todo, en tiempo, han quedado tareas pendientes para desarrollar en un futuro y que dependerá, en gran medida, del compromiso y confianza de la dirección en relación con la capacidad, demostrada en este trabajo, de detección de comportamientos anómalos y su vinculación con posibles infecciones, fallos hardware, fallos software o detección de ataques tanto pasados como en curso.

Entre las líneas de trabajo que han quedado pendientes y que podríamos consideras como continuación a la implementación ya obtenida tendríamos:

- ✓Aprovechar la implementación actual para avanzar en la detección de anomalías en comportamientos de servidores críticos que gestionen bases de datos, servicios web, servicios de correo, directorio activo o firewalls ampliando así la funcionalidad de detección a HIDS (Host Intrusion Detection System) mediante la instalación del plugging de métricas disponible en ELK Stack.
- ✓Crear más visualizaciones de todos los tipos disponibles para aprovechar toda la potencialidad que nos ofrece el entorno, ampliando la capacidad de detección de más comportamientos anómalos.
- ✓Crear más paneles o dashboards que integren de manera rápida visualizaciones relacionadas para la ayuda a la detección de anomalías.
- ✓Ampliar la funcionalidad de Zeek mediante la detección de amenazas conocidas basadas en firmas.
- ✓Ampliar las sondas, equipos con Zeek instalado, en el resto de centros hospitalarios de la empresa y en los segmentos de red que pueden afectar, en mayor medida, a la seguridad de servidores y aplicaciones críticas.

✓Activar la licencia empresarial de 30 días de prueba y valorar la propuesta a la dirección para su adquisición mediante la revisión de la funcionalidad completa X-PACK de:

- Funcionalidad avanzada Machine Learning
- Alerting
- Infraestructura
- Seguridad Endpoint
- Funcionalidades avanzadas de SIEM
- Etc.

✓Estudiar la posibilidad de contratación de Elastic Cloud, que nos permitiría disponer de un entorno con licencia de todos los productos identificados anteriormente pero disponiendo de una solución escalable cuyo crecimiento a futuro es incierto.

8. Glosario

- **Intrusión:** Una intrusión es una secuencia de acciones realizadas por un usuario o proceso deshonesto con el objetivo final de provocar un acceso no autorizado sobre un equipo o un sistema al completo.
- **Anomalía:** Existencia de una discrepancia de una regla o de un uso. Así pues, para poder detectar dicha discrepancia se tiene que definir primero lo que se considera como un comportamiento normal de un sistema. Una vez definido ésto, se realizará una clasificación como de sospechosa o intrusiva a aquellos comportamientos que se desvíen de lo que se considera como normal.
- **IDS**: Es una herramienta que tiene como función principal detectar intrusos en nuestros sistemas.
- **NIDS**: Sistema de detección de intrusiones en red.
- **HIDS:** Sistema de detección de intrusiones en host.
- **IPS:** Herramienta que permite la prevención de ataques.
- **SIEM:** Herramienta que nos permite recopilar, normalizar y correlacionar eventos de seguridad, proporciona inteligencia de seguridad, descarta falsos positivos, evalúa el impacto de un ataque, unifica la gestión de la seguridad, centraliza la información e integra herramientas de detección de amenazas.
- **Port Mirroring (puerto espejo):** Permite tomar el tráfico que atraviesa por un puerto, grupo de puertos o VLAN completa de un switch gestionado y enviar una copia hacia un puerto de destino definido.
- **Sniffer de Red:** Es un programa que captura paquetes de datos de la red.
- **Modo Promiscuo:** Para poder configurar un NIDS es necesario comprender el funcionamiento de una interfaz de red en "modo promiscuo" que es aquel en el que una computadora conectada a una red compartida, tanto la basada en cable de cobre como la basada en tecnología inalámbrica, captura todo el tráfico que circula por ella. Un NIDS se comporta realmente como un "sniffer de red" capturando todo el tráfico que en condiciones normales de funcionamiento se desecharía.
- **Antivirus**: Programas cuyo objetivo es detectar y eliminar virus informáticos.
- **Anti-malware**: Programa diseñado para prevenir, detectar y remediar software malicioso en los dispositivos informáticos individuales y sistemas TI. Puede incluir Anti-Spyware o detector de software espía.
- **Firewalls**: Elemento o software informático que trata de bloquear el acceso, a una red privada conectada a Internet, a usuarios no autorizados.
- **Parches de seguridad**: Actualizaciones de productos software para la eliminación de fallos/amenazas detectadas.
- **API REST**: Estándar lógico y eficiente para la creación de servicios web.
- **ACID**: Acrónimo de Atomicity, Consistency, Isolation and Durability: Atomicidad, Consistencia, Aislamiento y Durabilidad en español.

9. Bibliografía

9.1. Referencias bibliográficas:

[ABD15] Abdelkader Lahmadi, Frédéric Beck. «Powering Monitoring Analytics with ELK stack.» *9th International Conference on Autonomous Infrastructure, Management and Security.* Ghent, Belgium: hal-01212015, version 1, Junio 2015.

[ALO17] Alotaibi YK, Federico F. «The impact of health information technology on patient safety.» *Saudi Medical Journal*, Diciembre 2017: 1173-1180.

[AMR05] Amrit T. Williams, Mark Nicolett. *Improve IT Security With Vulnerability Management.* Gartner, 2005.

[AND72] Anderson, James, P. *Computer Security Technology Planning Study.* Hanscom Filed, Bedford, MA: ESD-TR-73-51, v II. Electronic Systems Division, Air Force Systems, Octubre 1972.

[BIN06] Bing, Chen, Lee J., y Wu A.S. *Fourth IEEE International Workshop on Information Assurance (IWIA'06).* London, UK, Abril 2006.

[COI19] Coimbra, Vinicius. *Data Hackers Medium.* s.f. https://medium.com/data-hackers/comparando-elasticsearch-vs-mongodb-4b5932c613d9 (último acceso: Octubre de 2019).

[DUS17] Mondek, Dusan, Rudolf B. Blažek, y Tomáš Zahradnický. «Security Analytics in the Big Data Era.» *2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C).* Prague, Czech Republic: DOI: 10.1109/QRS-C.2017.136, Julio 2017.

[GIM08] Giménez García, María Isabel. «Utilización de Sistemas de Detección de Intrusos como Elemento de Seguridad Perimetral.» Almería, 2008.

[SUN17] Son, Sung Jun, y Youngmi Kwon. «Performance of ELK stack and commercial system in security log analysis.» *2017 IEEE 13th Malaysia International Conference on Communications (MICC).* Johor Bahru, Malaysia: DOI: 10.1109/MICC.2017.8311756, Noviembre 2017.

[WYL12] Wylie, Brian, Daniel Dunlavy, Warren Davis, y Jeff Baumes. «Using NoSQL databases for streaming network analysis.» *DOI: 10.1109/LDAV.2012.6378986.* Seattle, WA, USA, Octubre 2012.

9.2. Referencias Web:

- ✓ https://www.sciencedirect.com/science/article/pii/S0716864017301268 --> Fecha 5/11/2019
- ✓ https://pdfs.semanticscholar.org/c3e2/1ac1621c8c9ccd89073bbca11543ba63020c.pdf --> Fecha 7/11/2019
- ✓ https://repositorio.unican.es/xmlui/handle/10902/4531 --> Fecha 12/11/2019
- ✓ https://protegermipc.net/2017/02/22/mejores-ids-opensource-deteccion-de-intrusiones/ --> Fecha 10/11/2019

- ✓ https://logz.io/blog/bro-elk-part-1/ --> Fecha 5/11/2019
- ✓ https://logz.io/blog/bro-elk-part-2/ --> Fecha 5/11/2019
- ✓ http://www.comparitech.com → Fecha 8/11/2019
- ✓ https://www.youtube.com/watch?v=-2MUKmndTAU --> Fecha 3/11/2019
- ✓ https://www.youtube.com/watch?v=HNVk31C2g4s--> Fecha 3/11/2019
- ✓ http://www.elastic.co --> Fecha desde el 01/11/2019 hasta el 10/12/2019
- ✓ http://www.zeek.org --> Fecha desde el 01/11/2019 hasta el 10/12/2019
- ✓ https://github.com/elastic/beats/pull/12812 --> Fecha 27/11/2019
- ✓ https://www.incibe.es/protege-tu-empresa/avisos-seguridad/importante-oleada-ransomware-afecta-multitud-equipos --> Fecha 08/12/2019
- ✓ https://www.avast.com/es-es/c-wannacry --> Fecha 09/12/2019
- ✓ https://blog.avast.com/es/what-is-bluekeep --> Fecha 09/12/2019
- ✓ https://www.welivesecurity.com/la-es/2019/11/08/crecen-ataques-bluekeep-windows/ -- > Fecha 12/11/2019
- ✓ https://www.who.int/patientsafety/es/ → Fecha 15/11/2019
- ✓ https://seis.es/wp-content/uploads/2018/02/Informe11.pdf --> Fecha 11/12/2019

9.3. Vídeos Visitados:

- ✓ https://www.youtube.com/watch?v=-2MUKmndTAU → Fecha 10/11/2019
- ✓ https://www.youtube.com/watch?v=HNVk31C2g4s → Fecha 10/11/2019

10. Anexo I - Características ELK Stack Basic

Características Incluidas en ELK Stack Basic (Suscripción gratis para siempre")

Tipos de almacenamiento

- Índice invertido (para búsqueda)
- Almacén de documentos (para documentos no estructurados)
- Almacén columnar (para analíticas)
- Árboles de BKD (para datos numéricos, geográficos y fechas)
- Tipo de campo aplanado
- Tipo de campo de forma
- Tipo de campo de vector
- Índices congelados (para almacenamiento a largo plazo)

Gestión de pilas

- Tutoriales de importación de datos
- Grok Debugger
- Upgrade Assistant (Asistente de actualización)
- Gestión de licencias

Clientes

- API REST
- Clientes de lenguaje
- Búsqueda en DSL
- Consola
- ES-Hadoop
- API y CLI de Elasticsearch SQL

Búsqueda de texto completo

- Relevancia
- Resaltado
- Autocompletar
- Correcciones
- Sugerencias
- Filtraciones
- Fijación de resultados
- Sinónimos actualizables dinámicamente
- Generador de perfiles de búsqueda

Gestión de datos

- API y UI de snapshot/restauración
- Snapshots mínimas
- Gestión de ciclo de vida de snapshot
- Data rollups
- Transformaciones de datos
- Gestión de indexación
- Gestión de ciclo de vida de indexación

Escalabilidad y resistencia

- Agrupación y alta disponibilidad
- Reequilibrio automático de datos
- Búsqueda entre clusters
- Nodos maestros solo de votación

Seguridad

- Comunicaciones encriptadas
- Control de acceso basado en roles
- Autenticación nativa y de archivos
- Kibana Spaces
- Controles de características de Kibana
- Gestión de claves de API

Monitoreo de pilas

- Monitoreo de pila completo

Analíticas

- Agregaciones
- Agregación de cardinalidad acumulativa

Machine Learning

- Visualizador de datos

Ingesta de productos y características

- Filebeat, Metricbeat, Winlogbeat, Packetbeat, Heartbeat, Auditbeat
- Functionbeat
- Logstash

- Funciones de similitud para campos de vectores

Fuentes de datos

- Sistemas operativos
- Servidores web y proxies
- Almacenes de datos y colas
- Servicios en la nube
- Contenedores y orquestación
- ArcSight CEF
- Datos del sistema de auditoría
- AWS
- AWS S3
- Azure
- Cisco ASA y Firepower
- CockroachDB
- CoreDNS
- Envoy Proxy
- Google Cloud Pub/Sub
- Flujos de VPC de Google Cloud
- Iptables
- Microsoft SQL Server
- NetFlow e IPFIX
- Base de datos Oracle
- Palo Alto PAN-OS
- Suricata
- Zeek (anteriormente Bro)

Canvas

- Canvas

Comparte y colabora

- Dashboards insertables
- UI y API de exportación de objetos
- Exportaciones de CSV
- Búsquedas guardadas

Logs de Elastic

- Shipper de log (Filebeat)
- Dashboards para fuentes de datos comunes
- App Logs
- Elastic Uptime, APM

Infraestructure

- ES-Hadoop
- Asistente de importación de archivos

Transformación de datos

- Enriquecimiento del tiempo de indexación
- Procesadores
- Analizadores
- Tokenizadores
- Filtros
- Grok
- Transformación de campo
- Enriquecimiento de búsqueda externa
- Procesador de ingesta de círculos

Elastic Common Schema

- Elastic Common Schema

Visualizaciones

- Series temporales
- Geo
- Métricas
- Tablas
- Nube de etiquetas
- Personalizado (Vega)

Exploración de datos

- Dashboards
- Discover
- Consola
- Característica de autocompletar búsquedas de Kibana

APM

- Servidor APM
- App APM
- Rastreo distribuido

Agentes APM

- Java
- .NET
- Go
- Ruby
- RUM (Javascript)
- Python

• Shipper de métrica (Metricbeat) • Dashboards para fuentes de datos comunes • App Infrastructure • Logs de Elastic, APM, tiempo de actividad ***Uptime*** • Monitor de tiempo de actividad (Heartbeat) • Dashboards de tiempo de actividad en Kibana • App Uptime • Logs, Infrastucture y APM de Elastic ***App Search*** • Servidor de App Search • UI de App Search • Curación de resultado de búsqueda • Analíticas de búsqueda • Sinónimos personalizados • Relevancia específica según el idioma • Modelo de relevancia con tolerancia a errores tipográficos • Ajuste del modelo de relevancia	• Nodo ***Integraciones*** • Logs e Infrastructure de Elastic ***SIEM*** • Elastic Common Schema • Análisis de seguridad del host • Análisis de seguridad de la red • Explorador de eventos de línea de tiempo • Integración con Maps ***Servicio Maps de Elastic*** • Mapas basados en capas • 18 niveles de zoom ***App Maps*** • Carga GeoJSON • Múltiples capas • Filtro basado en capas • Diseño del lado del cliente • Formas y puntos individuales • Agregaciones geográficas • Incorporación de mapas en el dashboard ***Seguridad*** • Comunicaciones encriptadas • Control de acceso basado en roles

11. Anexo II – Logs generados por Zeek

A continuación, se enumera los archivos de registro generados por Zeek, incluida una breve descripción del archivo de registro y enlaces a descripciones de los campos para cada tipo de registro extraído de la página oficial de Zeek.

Protocolos de Red

Log File	Description	Field Descriptions
conn.log	TCP/UDP/ICMP connections	**Conn::Info**
dce_rpc.log	Distributed Computing Environment/RPC	**DCE_RPC::Info**
dhcp.log	DHCP leases	**DHCP::Info**
dnp3.log	DNP3 requests and replies	**DNP3::Info**
dns.log	DNS activity	**DNS::Info**
ftp.log	FTP activity	**FTP::Info**
http.log	HTTP requests and replies	**HTTP::Info**
irc.log	IRC commands and responses	**IRC::Info**
kerberos.log	Kerberos	**KRB::Info**
modbus.log	Modbus commands and responses	**Modbus::Info**
modbus_register_change.log	Tracks changes to Modbus holding registers	**Modbus::MemmapInfo**
mysql.log	MySQL	**MySQL::Info**
ntlm.log	NT LAN Manager (NTLM)	**NTLM::Info**
radius.log	RADIUS authentication attempts	**RADIUS::Info**
rdp.log	RDP	**RDP::Info**
rfb.log	Remote Framebuffer (RFB)	**RFB::Info**
sip.log	SIP	**SIP::Info**
smb_cmd.log	SMB commands	**SMB::CmdInfo**
smb_files.log	SMB files	**SMB::FileInfo**

Log File	Description	Field Descriptions
smb_mapping.log	SMB trees	**SMB::TreeInfo**
smtp.log	SMTP transactions	**SMTP::Info**
snmp.log	SNMP messages	**SNMP::Info**
socks.log	SOCKS proxy requests	**SOCKS::Info**
ssh.log	SSH connections	**SSH::Info**
ssl.log	SSL/TLS handshake info	**SSL::Info**
syslog.log	Syslog messages	**Syslog::Info**
tunnel.log	Tunneling protocol events	**Tunnel::Info**

Ficheros

Log File	Description	Field Descriptions
files.log	File analysis results	**Files::Info**
ocsp.log	Online Certificate Status Protocol (OCSP). Only created if policy script is loaded.	**OCSP::Info**
pe.log	Portable Executable (PE)	**PE::Info**
x509.log	X.509 certificate info	**X509::Info**

Control de Red

Log File	Description	Field Descriptions
netcontrol.log	NetControl actions	**NetControl::Info**
netcontrol_drop.log	NetControl actions	**NetControl::DropInfo**
netcontrol_shunt.log	NetControl shunt actions	**NetControl::ShuntInfo**
netcontrol_catch_release.log	NetControl catch and release actions	**NetControl::CatchReleaseInfo**

Log File	Description	Field Descriptions
openflow.log	OpenFlow debug log	**OpenFlow::Info**

Detección

Log File	Description	Field Descriptions
intel.log	Intelligence data matches	**Intel::Info**
notice.log	Zeek notices	**Notice::Info**
notice_alarm.log	The alarm stream	**Notice::Info**
signatures.log	Signature matches	**Signatures::Info**
traceroute.log	Traceroute detection	**Traceroute::Info**

Observación de Red

Log File	Description	Field Descriptions
known_certs.log	SSL certificates	**Known::CertsInfo**
known_hosts.log	Hosts that have completed TCP handshakes	**Known::HostsInfo**
known_modbus.log	Modbus masters and slaves	**Known::ModbusInfo**
known_services.log	Services running on hosts	**Known::ServicesInfo**
software.log	Software being used on the network	**Software::Info**

Misceláneo

Log File	Description	Field Descriptions
barnyard2.log	Alerts received from Barnyard2	**Barnyard2::Info**
dpd.log	Dynamic protocol detection failures	**DPD::Info**
unified2.log	Interprets Snort's unified output	**Unified2::Info**

Log File	Description	Field Descriptions
weird.log	Unexpected network-level activity	**Weird::Info**
weird_stats.log	Statistics about unexpected activity	**WeirdStats::Info**

Diagnósticos de Zeek

Log File	Description	Field Descriptions
broker.log	**Peering status events between Zeek or Broker-enabled processes**	Broker::Info
capture_loss.log	**Packet loss rate**	CaptureLoss::Info
cluster.log	**Zeek cluster messages**	Cluster::Info
config.log	**Configuration option changes**	Config::Info
loaded_scripts.log	**Shows all scripts loaded by Zeek**	LoadedScripts::Info
packet_filter.log	**List packet filters that were applied**	PacketFilter::Info
prof.log	**Profiling statistics (to create this log, load policy/misc/profiling.zeek**	**N/A**
reporter.log	**Internal error/warning/info messages**	Reporter::Info
stats.log	**Memory/event/packet/lag statistics**	Stats::Info
stderr.log	**Captures standard error when Zeek is started from ZeekControl**	**N/A**
stdout.log	**Captures standard output when Zeek is started from ZeekControl**	**N/A**