Universitat Oberta de Catalunya
uoc.edu

Ventajas e Implementación de un sistema SIEM

ÁNGEL LUIS VELOY MORA
Máster Universitario en Seguridad de las Tecnologías de la Información y de las Comunicaciones
TFM-Seguridad empresarial

Jorge Chinea López
Víctor García Font

Diciembre de 2019

Dedicado a Maru y a Tere.

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Ventajas e Implementación de un Sistema SIEM*
Nombre del autor:	*Ángel Luis Veloy Mora*
Nombre del consultor/a:	*Jorge Chinea López*
Nombre del PRA:	*Víctor García Font*
Fecha de entrega (mm/aaaa):	12/2019
Titulación:	*Máster Universitario en Seguridad de las Tecnologías de la Información y de las Comunicaciones*
Área del Trabajo Final:	*TFM-Seguridad empresarial*
Idioma del trabajo:	*Castellano*
Palabras clave	*SIEM, Splunk, ELK*

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.*

El presente trabajo aborda dos objetivos diferentes. Por un lado, y a modo teórico, se demuestran todas las ventajas de implementar una solución SIEM dentro de una organización. A su vez, también se describe el estado actual de madurez de la última generación de SIEM que podemos encontrar dentro del mercado. Se demuestran las ventajas de juntar un SIEM, con un UBA/UEBA y un SOAR.

Por otro lado, de modo práctico, se implementa una solución híbrida de SIEM en la que todos los datos de la empresa son almacenados dentro del Data Lake (en este caso mediante ElasticSearch), para posteriormente enviar los eventos más destacados (eventos notables) a Splunk, para una vez almacenados en éste, implementar un SIEM con toda su inteligencia asociada.

Por tanto, el objetivo que se ha conseguido demostrar, ha sido que mediante una baja inversión económica, se puede implementar una solución SIEM totalmente funcional dentro de un SOC de una pequeña y mediana empresa.

Abstract (in English, 250 words or less):

The present paper solves two different objectives. In the first part, theoretically, we demonstrate all the advantages of implementing a SIEM solution within an organization. At the same time, we describe the current maturity status of the

latest generation of SIEM, that we can find within the trade. We have demonstrated the advantages of using a SIEM together witn an UEBA/UBA and a SOAR.

On the other hand, in a practical way, we have implemented a hybrid solution of SIEM in which all the company's data is stored within the Data Lake (in this case by ElasticSearch), to later send the most relevant events (notable events) to Splunk, for once stored in it, implement a SIEM with all its associated intelligence.

Therefore, it has been demonstrated that through a low monetary investment, a fully functional SIEM solution can be implemented within a SOC of a small and medium-sized company.

Índice

Lista de figuras

Lista de tablas

1. Introducción

1.1. Contexto y justificación del Trabajo

El presente TFM se desarrolla en el marco de una temática que está cada vez más de actualidad hoy en día. Los SIEM están cada vez más asentados dentro de los organigramas TIC de las empresas. Generalmente estos se encuentran dentro de un departamento SOC. Este servicio se puede dar tanto de forma interna en la empresa (on-premise), como subcontratándolo a un proveedor de servicios especializados (MSSP) en el ámbito de la seguridad de la información (sistema híbrido), o delegándolo totalmente (SIEM as a service).

El TFM cubre una doble necesidad. La primera de éstas cubre la necesidad de entender cuál es el estado del arte actual referente a los sistemas SIEM dentro de las organizaciones. Para ello se analizarán todas las funcionalidades que nos pueden proporcionar su uso, veremos los elementos constitutivos de estos, así como las distintas arquitecturas que nos podemos encontrar para desplegarlos. Se conseguirá por tanto demostrar las ventajas que supone implementar estos sistemas.

El segundo de los temas que abordaremos será dar una solución económica asumible por cualquier empresa a una de las funciones más importantes de un sistema SIEM y por ende de un SOC, que no es otro que la parte relacionada con la creación de las alertas o casos de uso. Este trabajo resuelve por tanto una problemática de índole económico que se puede dar dentro de una mediana o gran empresa que quiera poseer un Data Lake económico y que quiera a su vez poseer todas las ventajas que proporciona el sistema de correlación avanzado que se puede generar a través de un SIEM de última generación.

La solución al problema detectado se resuelve integrando ambos sistemas, el Data Lake representado en este caso por la solución Elastic Stack y que actualmente es totalmente gratuita (únicamente se paga si se desea tener soporte técnico) y donde se almacenarán cientos de gigas de datos diarios, y con una mínima inversión económica se contratará el servicio de Splunk Enterprise para dada su gran potencialidad, poder enviar a éste los logs más relevantes donde serán correlados de forma más efectiva. El Data Lake quedará por tanto para otras funciones como pueden ser el análisis forense de logs o el Threat Hunting.

Por tanto, lo que pretende este trabajo fin de máster es conseguir contextualizar el estado actual del arte dentro de los sistemas SIEM e implementar una solución económica para desplegar una alta capacidad de correlación para la posterior generación de casos de uso más eficientes.

1.2. Objetivos del Trabajo

En el presente trabajo podemos apreciar dos objetivos bien diferenciados. Un teórico y otro práctico.

El primer objetivo que es teórico se encarga de realizar una descripción detallada de todo lo relacionado con los sistemas SIEM actuales. En esta sección se analizarán conceptos como pueden ser su motivación, su funcionalidad hasta las posibles arquitecturas existentes. Con este estudio se pretende lograr el objetivo de demostrar todas las ventajas a nivel de seguridad que supone la implementación de un SIEM dentro de una organización.

La segunda parte eminentemente práctica, está relacionada con la generación dentro de un Data Lake como es Elastic Stack de una serie de eventos notables o eventos destacados que, ayudándonos de una serie de alertas generadas dentro del propio Elastic, integrar los datos contenidos dentro del Data Lake con la infraestructura del Splunk Enterprise. Dentro de este último sistema se generará un sistema de correlación complejo usando los eventos provenientes del Data Lake para crear una serie de alertas o casos de uso.

1.3. Enfoque y método seguido

En referencia a la parte teórica, el enfoque y método usado será el estudio y análisis de la documentación especializada existente generada tanto por los fabricantes como por investigadores a través de documentos técnicos. Una vez realizado el análisis en detalle seremos capaces de enumerar y detallar cada uno de los componentes constitutivos actuales de los sistemas SIEM, y su importancia para la organización donde se implementa.

El enfoque práctico del proyecto es conseguir integrar las dos soluciones SIEM, de forma que por un lado tengamos un almacén muy grande y barato de datos (Data Lake) y por otro tengamos un correlador de alertas altamente flexible y potente.

Por todo ello, la segunda parta eminentemente práctica, consistirá por un lado en desarrollar un proceso de análisis de los elementos constitutivos principales de ambas arquitecturas a usar, tanto a nivel de Elastic Stack y Splunk, y por otro, la implementación de la interconexión entre los dos sistemas y la generación de los elementos de correlación.

Se analizarán los dos correladores de eventos en origen (ElastAlert y Watcher), de las posibilidades a nivel de correlación que nos proporciona la solución Splunk Enterprise gracias a la gran potencialidad de su lenguaje de consulta (SPL) para poder generar de forma eficaz las distintas alertas, así como de todos los elementos teóricos necesarios para generar conceptos tales como los de eventos notables, alertas de alto nivel o cómo interactúan dichos elementos entre sí.

Dentro de las posibilidades que había para poder enviar los eventos notables desde el Elastic a Splunk, se ha elegido el realizarlo a través de un método POST dado que se ha determinado que es un método sencillo de implementar y que nos permite enviar los logs de la forma personalizada deseada. Esto es muy importante dado que será en origen donde se generará la normalización

de los eventos y la creación de nuevos campos en forma de etiquetas que posteriormente nos ayudará en la correcta generación de los casos de uso.

1.4. Planificación del Trabajo

Los recursos para realizar el trabajo son los siguientes:

- **Arquitectura Data Lake**. Se trata de toda la arquitectura relacionada con Elastic Stack implementada de forma distribuida para hacer las veces de lago de datos de la empresa.

  Dentro de ésta se incluyen los dos correladores que generarán los eventos destacados de seguridad. Destacar que el aplicativo del ElastAlert se encontrará en un servidor no dedicado en alta disponibilidad (Watcher se integra dentro de la arquitectura de Elastic Stack).
- **Arquitectura Splunk Enterprise**. Es la arquitectura desplegada de forma distribuida relacionada con el SIEM.
- **Feeds externos de seguridad** que nos proporcionarán el enriquecimiento de los eventos y alertas.
- **Dispositivos origen de los eventos** a monitorizar más aquellos que permiten interconectar todos los sistemas existentes dentro del proyecto:

  - Servidores y Workstation generadores de eventos.
  - Elementos de red. Como pueden ser los FWs de los cuales obtendremos los eventos de red de la empresa.
  - Otros dispositivos de seguridad como pueden ser IDS, EDR, appliance de seguridad de correo, …
  - Y por último todo elemento a nivel de administración y comunicaciones para hacer posible la implementación de todo el ecosistema a monitorizar y que posibiliten las comunicaciones.
- **Acceso a recurso en red o físicos** para la investigación de las dos partes del proyecto.
- **Ordenador personal y toda la paquetería ofimática** para generar tanto la memoria como la presentación y defensa del proyecto.
- **Recursos humanos**. Investigador creador del TFM que debe poseer amplios conocimientos y experiencia en el campo de la seguridad de la información y en la integración de dispositivos de seguridad.

Las tareas a realizar son las siguientes:

<u>Parte teórica.</u>

- Ventajas de los sistemas SIEM.
- Componentes y capacidades de los SIEM.
- Funcionamiento de los SIEM.

- Funciones de los SIEM.
- Evolución de los sistemas SIEM.
- Buenas prácticas.
- Arquitectura de los SIEM. Arquitectura lógica. Arquitectura física.
- Agregación, procesamiento y análisis de logs de seguridad.
- User and Entity Behavior Analytics (UEBA).
- Análisis de la seguridad a través del Big Data.
- Orquestación de la seguridad y automatización de la respuesta ante incidentes mediante SOAR.
- El SOC y la integración con el SIEM.
- Comparativa SIEM.
- Búsqueda de amenazas (Threat Hunting) y los sistemas SIEM.

Parte práctica.

- Analizar las capacidades que nos provee tanto ElastAlert como Watcher para generar como resultado del trigger de una alerta una acción de tipo POST enviada a Splunk.
- Analizar la forma de recepcionar los eventos provenientes del Elastic Stack dentro de Splunk. Crear el índice apropiado y comprobar mediante una prueba la correcta recepción e indexación.
- Determinar qué assets serán los que vamos a monitorizar a través de Splunk.
- Una vez determinado lo anterior, detectar y filtrar todos los eventos relevantes de seguridad que generaremos de forma directa o agregada a través de los correladores del Elastic Stack.
- Una vez tengamos los eventos notables, normalizar todos los campos para facilitar su posterior correlación.
- Generar las etiquetas de enriquecimiento de los eventos que se van a enviar a Splunk.
- Una vez que estamos recibiendo los eventos dentro de Splunk, enriquecer los datos mediante CSV de información interna e información de fuentes externas.
- Crear los casos de uso.

Planificación temporal.

Hay que indicar que, dado que la propia naturaleza del proyecto lo permite, se va a desarrollar los dos temas principales a la vez. Esto se verá reflejado en la planificación temporal del proyecto.

Los plazos del TFM son marcados por las propias entregas planificadas dentro del Aula Virtual. Las entregas son las siguientes:

Entrega 1 – Plan de trabajo - 18/09/19 a 01/10/19
Entrega 2 - 02/10/19 a 29/10/19
Entrega 3 - 30/10/19 a 26/11/19
Entrega 4 – Memoria final - 27/11/19 a 31/12/19
Entrega 5 – Presentación en vídeo - 01/01/20 a 07/01/20
Defensa TFM - 13/01/20 a 17/01/20

Tabla 1: Planificación temporal hitos TFM

Por tanto, la planificación temporal en base a lo anterior y considerando que la carga crediticia del TFM es de 25 horas por crédito, hay que cubrir un total de 225 horas.

Actividad	Fecha inicio	Fecha fin	Duración
Planificación del TFM	18/09/19	30/09/19	12
Elaboración parte teórica TFM	02/10/19	29/10/19	60
Elaboración parte práctica TFM	30/10/19	31/12/19	132
Preparación del vídeo de la defensa	01/01/20	07/01/20	15
Preparación de la defensa del TFM	07/01/20	17/01/20	6
			225

Tabla 2: Resumen planificación TFM

Actividad	Fecha inicio	Fecha fin	Duración
Planificación del TFM			
Determinación del plan de trabajo a realizar con el tutor	18/09/19	26/09/19	2
Elaboración tema primero	27/09/19	30/09/19	10
Entrega 1 – Plan de trabajo	01/10/19	01/10/19	Hito
Elaboración parte teórica del TFM			
Ventajas de los sistemas SIEM	02/10/19	05/10/19	5
Componentes y capacidades de los SIEM	06/10/19	12/10/19	4
Funcionamiento de los SIEM	13/10/19	14/10/19	4
Funciones de los SIEM	15/10/19	18/10/19	4
Evolución de los sistemas SIEM	19/10/19	19/10/19	1
Buenas prácticas	20/10/19	20/10/19	3
Arquitectura de los SIEM. Arquitectura lógica. Arquitectura física	21/10/19	22/10/19	9
Agregación, procesamiento y análisis de logs de seguridad	22/10/19	23/10/19	5
User and Entity Behavior Analytics (UEBA)	24/10/19	25/10/19	5
Análisis de la seguridad a través del	25/10/19	25/10/19	2

Big Data			
Orquestación de la seguridad y automatización de la respuesta ante incidentes mediante SOAR	26/10/19	27/10/19	5
El SOC y la integración con el SIEM	27/10/19	27/10/19	3
Comparativa SIEM	28/10/19	28/10/19	5
Búsqueda de amenazas (Threat Hunting) y los sistemas SIEM	28/10/19	29/10/19	5
Entrega 2	29/10/19	29/10/19	Hito
Elaboración parte práctica del TFM			
Análisis de las capacidades del ELK	30/10/19	7/11/19	12
Análisis de las capacidades de Splunk	8/11/19	14/11/19	14
Determinación de fuentes a integrar	15/11/19	17/11/19	8
Generar los eventos notables	18/11/19	25/11/19	18
Entrega 3	26/11/19	26/11/19	Hito
Normalización de los eventos notables	27/11/19	08/12/19	25
Creación de los casos de uso	09/12/19	20/12/19	25
Redacción y corrección de la memoria a entregar	21/12/19	30/12/19	30
Entrega 4 – Memoria final	31/12/19	31/12/19	Hito
Preparación del vídeo de la defensa			15
Entrega 5 – Presentación en vídeo	07/01/20	07/01/20	Hito
Preparación de la defensa del TFM			6
Defensa TFM	17/01/20	17/01/20	Hito

Tabla 3: Planificación detallada TFM

A continuación, se muestra el diagrama de Gantt de la planificación temporal del trabajo fin de máster.

	Nombre	Inicio	Terminado
1	**Planificación del TFM**	**18/09/19 8:00**	**1/10/19 16:00**
2	Determinación del plan de trabaj...	18/09/19 8:00	26/09/19 17:00
3	Elaboración tema primero	27/09/19 8:00	30/09/19 17:00
4	Entrega 1	1/10/19 8:00	1/10/19 16:00
5	**Elaboración parte teórica del T...**	**2/10/19 8:00**	**29/10/19 17:00**
6	Ventajas de los sistemas SIEM	2/10/19 8:00	5/10/19 17:00
7	Componentes y capacidades de l...	6/10/19 8:00	12/10/19 17:00
8	Funcionamiento de los SIEM	13/10/19 8:00	14/10/19 17:00
9	Funciones de los SIEM	15/10/19 8:00	18/10/19 17:00
10	Evolución de los sistemas SIEM	19/10/19 8:00	19/10/19 16:00
11	Buenas prácticas	20/10/19 8:00	20/10/19 16:00
12	Arquitectura de los SIEM. Arquite...	21/10/19 8:00	22/10/19 17:00
13	Agregación, procesamiento y an...	22/10/19 8:00	23/10/19 17:00
14	User and Entity Behavior Analytic...	24/10/19 8:00	25/10/19 17:00
15	Análisis de la seguridad a través ...	25/10/19 8:00	25/10/19 16:00
16	Orquestación de la seguridad y a...	26/10/19 8:00	27/10/19 17:00
17	El SOC y la integración con el SIEM	27/10/19 9:00	27/10/19 17:00
18	Comparativa SIEM	28/10/19 9:00	28/10/19 17:00
19	Búsqueda de amenazas (Threat ...	29/10/19 9:00	29/10/19 17:00
20	Entrega 2	29/10/19 9:00	29/10/19 17:00
21	**Elaboración parte práctica del ...**	**30/10/19 9:00**	**20/12/19 17:00**
22	Análisis de las capacidades del ELK	30/10/19 9:00	7/11/19 17:00
23	Análisis de las capacidades de Spl...	8/11/19 9:00	14/11/19 17:00
24	Determinación de fuentes a integ...	15/11/19 9:00	17/11/19 17:00
25	Generar los eventos notables	18/11/19 9:00	25/11/19 17:00
26	Entrega 3	26/11/19 9:00	26/11/19 17:00
27	Normalización de los eventos not...	27/11/19 9:00	8/12/19 17:00
28	Creación de los casos de uso	9/12/19 9:00	20/12/19 17:00
29	Redacción y corrección de la memo...	21/12/19 9:00	30/12/19 17:00
30	Entrega 4	31/12/19 9:00	31/12/19 17:00
31	Preparación del vídeo de la defensa	1/01/20 9:00	6/01/20 17:00
32	Entrega 5	7/01/20 9:00	7/01/20 17:00
33	Preparación de la defensa del TFM	8/01/20 9:00	16/01/20 17:00
34	Defensa TFM	17/01/20 9:00	17/01/20 17:00

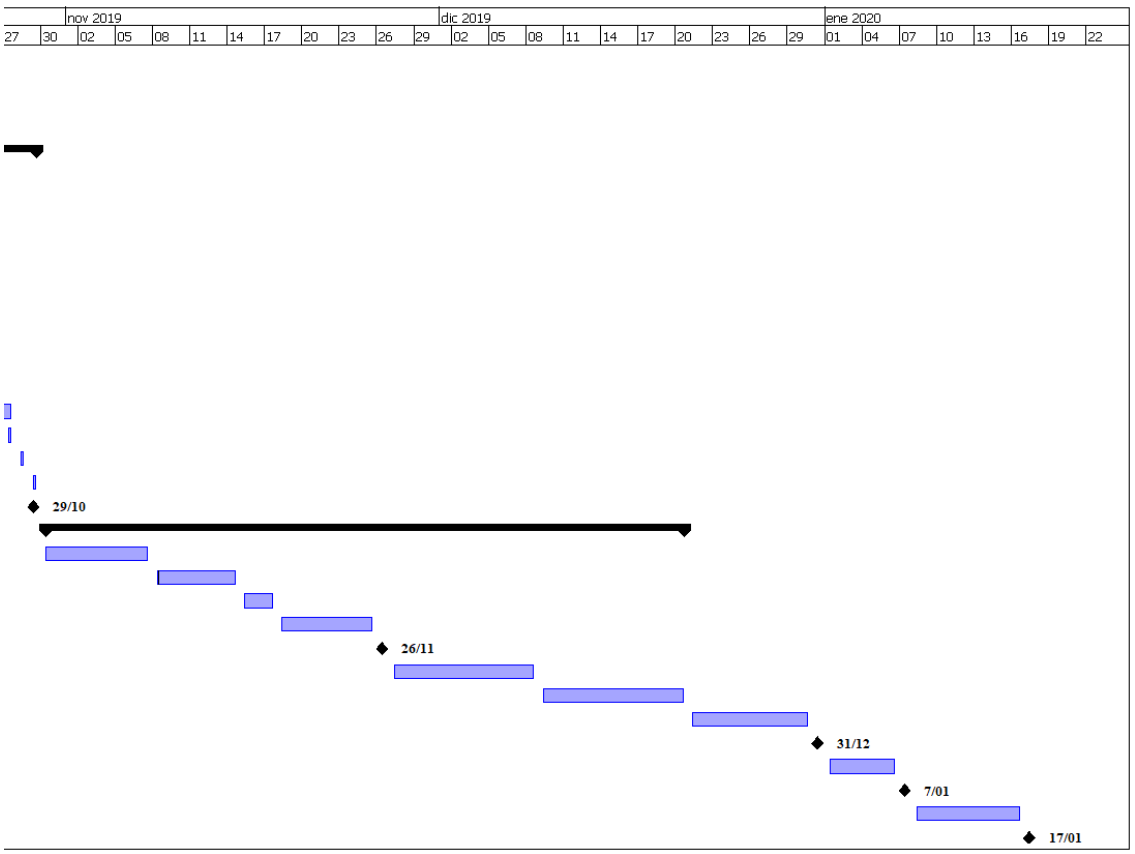

Ilustración 1: Diagrama de Gantt TFM

1.5. Breve sumario de productos obtenidos

Los productos obtenidos están relacionados con las distintas PEC que consta el presente TFM. Se entregarán al finalizar el trabajo, y constan de la memoria final del proyecto, junto con el vídeo de la defensa.

1.6. Breve descripción de los otros capítulos de la memoria

Como se ha podido ver el trabajo consta de los siguientes capítulos:

- **Introducción**. Capítulo en el que se listan los objetivos y la forma de conseguirlos referentes a la realización del TFM.
- **Ventajas de la implementación de un sistema SIEM**. El segundo capítulo versa sobre las ventajas del uso de un sistema SIEM. Capítulo teórico relacionado con todo lo que supone un SIEM.
- **Implementación SIEM**. El tercer capítulo principalmente práctico, se centra en la implementación de un correlador avanzado para la creación de casos de uso.
- **Conclusiones finales**. El cuarto y último capítulo está relacionado con las conclusiones finales y líneas a desarrollar en el futuro.

1.7. Estado del arte

El estado actual del arte referente a los SIEM es que nos encontramos enmarcados dentro de la que se considera la tercera generación de estos sistemas. Los SIEM actuales enmarcados dentro de los SOC, son capaces no solo de monitorizar eventos de seguridad en tiempo real, generar alertas y reportes para su uso por los miembros del equipo de seguridad, sino que actualmente y gracias a elementos complementarios como los UEBA y SOAR, son capaces de realizar funcionalidades avanzadas dentro del campo de la seguridad como puede ser la detección de exfiltración de datos, detección amenazas de tipo Zero-day, detección de actividades sospechosas que pueden ser indicios de algún tipo de incidente de seguridad, … Además, permiten a los analistas de seguridad realizar Threat Hunting con mayor éxito. Por último, y gracias a la integración con los SOAR, son capaces de responder ante las amenazas de una manera rápida, coordinada y automática.

En cuanto al propósito de la parte práctica, actualmente no existe una forma estándar y mucho menos automatizada para realizar la integración de los logs más destacados (eventos notables) contenidos dentro de un Data Lake con un SIEM comercial que no lo haga de forma nativa. Existen formas de enviar los logs desde el ELK a Splunk, pero no de manera tan selectiva y enriquecida como de la forma en que en el presente trabajo se propone.

2. Ventajas de la implementación de un sistema SIEM

En el presente capítulo se va a describir la parte teórica de las ventajas de desplegar una solución SIEM con respecto a los modelos anteriores. A lo largo del capítulo vamos a ir viendo los distintos componentes de los que constan los SIEM modernos y qué ventajas producen en las organizaciones donde son desplegados.

Lo primero que debemos realizar es la definición formal de lo que es un SIEM. Para ello nos vamos a basar en la propia definición que le da Gartner:

"*La tecnología SIEM admite la detección de amenazas y la respuesta ante incidentes de seguridad, a través de la recopilación en tiempo real y el análisis histórico de eventos de seguridad de una amplia variedad de fuentes de datos contextuales y de eventos. También admite informes de cumplimiento e investigación de incidentes a través del análisis de los datos históricos provenientes de estas fuentes. Las capacidades centrales de la tecnología SIEM son un amplio alcance de recopilación de eventos y la capacidad de correlacionar y analizar eventos de fuentes dispares*".

El término SIEM proviene de la unión de dos tipos distintos de tecnologías como son el SIM (Security Information Management) y el SEM (Security Event Management). Como se puede deducir en base a su nombre, el primero se encarga de la administración de la Seguridad de la Información, y el segundo de la administración de los eventos de seguridad. Por todo ello, ya podemos hacernos una idea general sobre en qué va a consistir a nivel funcional un SIEM. Dado que dichas tecnologías son fundamentales, vamos a profundizar un poco más en sus características. Antes que nada, indicar que ambos tipos de sistemas ya almacenan los logs de manera centralizada. Con toda esta información, ahora sí pasemos a verlos en detalle:

- **SIM**. Considera como la primera generación, hace referencia a la recolección de los logs de los sistemas para su administración. Con el SIM se introdujo conceptos tan importantes como el almacenaje de logs durante largos períodos de tiempo, el realizar análisis sobre estos y la creación de reportes. Además, introduce la posibilidad de combinar los logs recopilados con fuentes externas de Threat Intelligence para su enriquecimiento.

- **SEM**. Este término hace referencia a una segunda generación en la que su objetivo es centrarse en los eventos de seguridad. Para ello se realizan con estos agregaciones, correlaciones y notificaciones con todos aquellos eventos provenientes de las distintas fuentes que tengan relevancia en el ámbito de la seguridad. El SEM con sus datos centralizados permite un análisis de eventos de seguridad en casi tiempo real.

A modo de resumen podemos decir que el SEM destaca por su potencia en la monitorización en tiempo real, potenciándose las notificaciones, mientras que el SIM destaca por su almacenaje a largo plazo y por tanto por su capacidad de analizar los eventos con mayor profundidad.

El contexto de la necesidad de implementar un SIEM puede ser visto desde varios puntos. En primer lugar, las empresas hoy en día están cada vez más digitalizadas. Esto hace que la cantidad de datos que se generen se haya incrementado de manera drástica. Por otro lado, el tener más y más sistemas interconectados y expuestos, ha generado que la superficie de ataque de las organizaciones se haya incrementado hasta el punto de que el riesgo de sufrir un ataque sea algo totalmente real. Máxime cuando los ataques relacionados con la ciberseguridad se han convertido en un negocio profesional para las organizaciones criminales, como un objetivo estratégico para gobiernos y organizaciones competidoras entre sí. Sin olvidar las ya clásicas motivaciones de hacer daño por hacer daño, o las amenazas internas que pueden suponer empleados descontentos.

Todo ello hace que ante un incidente de seguridad se deba ser lo más rápido y eficaz posible. Si todos estos logs generados por los distintos sistemas están deslocalizados de forma local, la detección, investigación, análisis, … de dichos incidentes se hacen casi imposible de detectar. De este escenario surge la necesidad de por un lado centralizar todos estos logs generados por los sistemas, y por otro, poder gestionarlos de forma eficiente y aplicar sobre estos acciones que ayuden a detectar posibles incidentes de seguridad.

En referencia a cómo se actuaba, en un principio surgió la acción reactiva ante los distintos eventos que iban generando los SIEM. Se detectaba un incidente y se actuaba en consecuencia. Hoy en día a esta medida se ha añadido la posibilidad de implementar acciones proactivas para llegado el caso poder prevenir dichos ataques o detectarlos en sus fases iniciales, reduciendo así su impacto. Para lograr esto último, se han ido añadiendo a los SIEM tradicionales nuevos módulos que en un principio eran independientes como son los UEBA o los SOAR, pero que juntos consiguen una sinergia y unos resultados mucho mejores ante la gestión de incidentes de seguridad.

2.1. Ventajas de los sistemas SIEM.

A continuación, vamos a enumerar todas las ventajas existentes que nos da la implementación de un sistema SIEM dentro de una organización.

- Proactividad relacionada con los incidentes de seguridad.
- Rapidez de respuesta ante incidentes detectados.
- Detección de amenazas previamente desconocidas. Esto se puede lograr mediante el uso de analítica avanzada de eventos.
- Mayor velocidad a la hora de llevar a cabo la investigación de las alertas generadas.

- Posibilidad de detectar amenazas en logs históricos. Para este objetivo emplea herramientas de analítica.
- Monitorización de las actividades que se llevan a cabo dentro de la red. Actividad de los usuarios y los dispositivos administrados.
- Garantiza la protección de los datos. Cumplimiento de normativa y legislación. Mejora de las operaciones de la empresa.
- Detección de activos. Evalúa todos los activos de la red. Para ello se basa en el escaneo activo de red, monitoreo pasivo, inventario de activos, inventario de software.
- Evaluación de vulnerabilidades. Identifica las vulnerabilidades de los sistemas administrados, realiza pruebas de vulnerabilidades a nivel de red y monitoriza de manera continua para la detección de nuevas vulnerabilidades.

2.2. Componentes y capacidades de los SIEM.

A continuación, pasamos a describir los componentes y capacidades que poseen los sistemas SIEM hoy en día.

- **Agregación de datos**. Función básica por la que se agregan los eventos de seguridad que posteriormente serán la base que se usará en los demás puntos. Estos eventos provienen de múltiples fuentes y son de vital importancia el determinar su relevancia y si se deben indexar o no.

- **Correlación**. Este término hace referencia a la acción de vincular eventos y datos relacionados entre sí en un conjunto que puede representar un incidente de seguridad real, una amenaza, una vulnerabilidad o un evento de tipo forense. Básico para el éxito de la correlación es la capacidad del propio SIEM a la hora de realizar consultas. Por ejemplo, el lenguaje de Splunk conocido como SPL es uno de los más potentes del mercado, lo que hace que se puedan sacar más partido a las correlaciones y por tanto a los resultados devueltos.

- **Analítica**. La analítica hace uso de modelos estadísticos y aprendizaje automático (machine learning) para identificar relaciones más profundas (que en un primer momento no serían fácilmente detectables) entre los datos y anomalías detectadas frente a tendencias conocidas o habituales. Con dicha analítica se pueden llegar a detectar amenazas que de otro modo no podrían haber sido detectadas. En los últimos tiempos se han introducido dentro de los SIEM para este fin los módulos denominados UEBA (User and Entity Behaviour Analytics) / UBA (User and Behaviour Analytics).

- **Uso de fuentes externas de tipo Threat Intelligence**. Se trata de combinar los datos internos con fuentes externas de Threat Intelligence. Éstas últimas contienen datos relacionados con vulnerabilidades, patrones de ataque u otros indicadores maliciosos. A estos datos se

conocen comúnmente como IOCs. Una de las plataformas más conocida es el proyecto MISP (Malware Information Sharing Platform).

- **Alertas**. Se trata de la acción de analizar eventos en busca de alguna condición de disparo. Una vez que ésta se cumple, alerta al equipo de seguridad mediante algún tipo de acción como puede ser el envío de un correo u otro mensaje o mediante la aparición de dicho evento dentro de algún panel de monitorización especialmente dedicado para ello.

- **Dashboards y visualizaciones**. Es la capacidad de crear visualizaciones en base a los datos para facilitar la revisión de estos e identificar actividades que no se ajustan a patrones estándar. Son menos potentes que la analítica, pero ante picos extraños pueden ayudar para una rápida detección.

- **Compliance**. Automatiza la recopilación de datos relacionados con el cumplimiento normativo de una empresa, produciendo informes que se adaptan a los procesos de seguridad, gobierno y auditoría para una serie de estándares como puede ser PCI/DSS.

- **Retención**. Es la función del almacenaje de los datos a lo largo del tiempo. Esto hace que se puedan realizar análisis más completos, seguimientos de ciertas actividades y cumplir con ciertas políticas de retención de logs. La política de retención puede ser por ejemplo en base al tiempo o en base al tamaño que ocupa almacenado.

- **Threat Hunting**. Función primordial dentro de cualquier SOC maduro, se trata del análisis en base a consultas, filtrados o pivoteo de los datos para descubrir amenazas o vulnerabilidades de forma proactiva. Esto se basa en la premisa de que un compromiso no tiene porqué ser detectado por los distintos appliances de seguridad existentes o por las reglas correladas reactivas implementadas dentro del SIEM. En base a por ejemplo los TTPs se pueden detectar indicadores en principio ocultos dentro de los eventos almacenados en los SIEM. Un claro ejemplo de qué se podría detectar a través de esta metodología sería la compleja detección de una APT.

- **Incident Response**. Se trata de otra de las funciones más importantes a implementar dentro de los SIEM actuales. Se refiere a las acciones que se deben tomar ante la detección de un incidente. Los SIEM actuales proporcionan la gestión de los incidentes, colaboración e intercambio de conocimientos sobre éstos. Todo ello hace que la gestión de los incidentes sea más ágil y eficiente, dado que facilita la relación con otros incidentes iguales o relacionados a lo largo del tiempo.

- **Automatización SOC**. Otra de las grandes características de los SIEM de tercera generación. Se denominan SOAR (Security Orchestration, Automation and Response) y es la parte del SIEM que se integra a través de las APIs con otros appliances de seguridad para crear playbooks y workflows automatizados que son disparados ante una serie

de eventos específicos. Gracias a esto, se reducen drásticamente los tiempos de respuesta ante un incidente dado, y siempre gestionados en base a unos procedimientos estándares y bien definidos.

2.3. Funcionamiento de los SIEM.

El funcionamiento de un sistema SIEM se define básicamente como las acciones que puede realizar dentro un correcto funcionamiento.

- **Recolección de datos**. Los SIEM para la recolección de los datos suelen desplegar una serie de agentes en los dispositivos finales, servidores, equipos de red u otros sistemas. La otra opción que se suele usar es el uso de protocolos de reenvío de syslog, SNMP o WMI. Pueden a su vez integrar otros tipos de fuentes de datos no tan estándares o incluso integrarse con los servicios en la nube para la obtención de datos de registro sobre la infraestructura implementada en ésta o de aplicaciones SaaS. El procesamiento que pueden darse a los eventos puede ocurrir en estos elementos alojados en el dispositivo final. Se pueden realizar sobre todas o solo algunas fuentes antes de su envío para su almacenamiento centralizado.

- **Almacenamiento de datos**. Tradicionalmente los SIEM han estado almacenando sus datos en los data center tradicionales de la empresa. Esto hacía que se dificultase el propio almacenamiento y administración de los grandes volúmenes de datos que pueden llegar a tener dichos sistemas. Todo esto hacía que únicamente se almacenaran unos pocos registros. Los SIEM actuales se basan en tecnologías nuevas como son los Data Lake que posibilitan escalabilidad casi ilimitada de almacenamiento a bajo coste. Esto hace que se permitan retener y analizar todos los eventos almacenados provenientes de múltiples fuentes.

- **Políticas y reglas**. Una de las capacidades de los SIEM es que se permite definir cómo se comportan los sistemas a monitorizar. Una vez establecido lo que es normal, se establecen reglas y umbrales que definirán qué anomalías pueden llegar a considerarse como un incidente de seguridad. Esta capacidad que en un principio es manual y de difícil implementación dado que deben de conocerse profundamente los sistemas y que suelen generar falsos positivos a la vez que se ocultan anomalías sutiles ha ido sustituyéndose a lo largo del tiempo por el uso del aprendizaje automatizado para detectar automáticamente anomalías y definir de forma autónoma reglas sobre los datos, todo ello hace que se lleguen a descubrir eventos de seguridad difícilmente detectables para su investigación.

- **Consolidación y correlación de datos**. Un propósito central de los sistemas SIEM es la recolección de todos los datos que sean necesarios monitorizar y sobre estos generar las distintas correlaciones. Correlar no es más que unir eventos relacionados entre sí a lo largo de una línea temporal que pueden llegar a ser significativos en el ámbito de la

seguridad. El resultado de dicha correlación está destinado a que sobre éste se produzca algún tipo de acción.

2.4. Funciones de los SIEM.

A continuación, vamos a enumerar cuáles son las funciones de los SIEM.

- **Monitorizaciones de seguridad**. Los SIEM se encargan de monitorizar en tiempo real los sistemas de la organización para detectar posibles incidentes de seguridad. La ventaja del SIEM es que centraliza todos eventos relevantes para la seguridad de la empresa. Sin él, tendríamos los eventos distribuidos en múltiples dispositivos no relacionados entre ellos que dificultarían en extremo su correlación y por tanto detección y puesta en marcha de una respuesta. El SIEM interrelaciona contextualizando a lo largo de una línea temporal un conjunto de eventos que ahora sí tiene un sentido dentro de la seguridad.

- **Detección avanzada de amenazas**. Los SIEM ayudan a detectar, mitigar y prevenir amenazas avanzadas. Por ejemplo, con un SIEM podemos detectar los denominados insiders, mediante el análisis de todos los eventos que generan los usuarios dentro de la organización. Se pueden usar también para la detección exfiltración de datos. La exfiltración no es otra cosa que fuga de información realizada de forma ilícita fuera de la organización. Se puede detectar por eventos de red en base a anomalías, o bien integrando dentro del SIEM los logs de dispositivos especializados como son los DLP (Data Loss Prevention). Incluso un SIEM puede llegar a detectar en base a sus capacidades una de las amenazas más difíciles de detectar y que son de las más peligrosas, nos estamos refiriendo a las APTs (Advanced Persistent Threat).

- **Análisis forense y respuesta ante incidentes de seguridad**. Los analistas, haciendo uso de los SIEM poseen la capacidad de detectar incidentes de seguridad que se están produciendo en ese preciso momento, ayuda a poder clasificarlo y a poder dar los pasos necesarios para implementar una remediación. El detectar un incidente supone el tener que determinar si anteriormente ya existían indicios de compromiso. Para ello se realiza el llamado análisis forense de los logs para determinar a lo largo de un timeline el inicio de dicho incidente y los eventos sucesivos que se han registrado hasta el momento presente. Básicamente se busca detectar las distintas fases del ataque (un ejemplo de metodología que se suele seguir es el Cyber Kill Chain de Lockheed Martin) que ha generado el supuesto incidente. Esta recopilación y análisis de los datos es costosa en tiempo y recursos, los SIEM facilitan toda esta labor al ya poseer todos los logs necesarios almacenados y poseer herramientas de búsqueda avanzadas y funciones de pivoting. Por tanto, un analista no solamente se enfrenta a detecciones en tiempo real, sino que pueden detectar amenazas pasadas mediante técnicas de investigación. Como resultado tendremos los suficientes datos como para determinar qué ha pasado, quién estaba

implicado y qué decisiones se deben tomar para la recuperación del incidente.

- **Revisión y creación de reportes relacionados con el Compliance de las políticas de seguridad**. Los SIEM ayudan a las organizaciones a demostrar a los auditores y reguladores que cuentan con las garantías adecuadas y que los incidentes de seguridad son conocidos y están contenidos. Desde los orígenes han sido uno de los usos dados a los SIEM. Actualmente muchos de estos ya cuentan con procesos de monitorización automática y con la generación de informes para el cumplimento de estándares como PCI/DSS, HIPPA, SOX, FERPA e HITECH.

2.5. Evolución de los sistemas SIEM.

Veamos cómo los SIEM han ido evolucionando a lo largo del tiempo.

- **1ª Generación**. La primera generación de SIEM combinaba un sistema SIM con un SEM. Estaban limitados en cuanto al escalado de los datos que se podían administrar. Por último, soportaban alertas y visualizaciones.

    - Escalabilidad: Verticalmente.
    - Retención de datos: Parcial.
    - Recolección de datos: Ingesta lenta y manual de datos.
    - Detección de amenazas: Análisis manual y alertas basadas en reglas manuales.
    - Respuesta ante incidentes: Poca o nula.
    - Dashboards y visualizaciones: Muy limitadas.

- **2ª Generación**. Esta segunda generación introduce la infraestructura de Big Data. De forma centralizada se permite gestionar eventos almacenados durante períodos largos de tiempo, usar el concepto de consulta en tiempo real e implementar el Threat Intelligence. Todo ello permite una visión integral de los datos de la empresa.

    - Escalabilidad: Horizontalmente, soportando Big Data.
    - Retención de datos: Completo. Permite filtrado.
    - Recolección de datos: Ingesta automatizada. La fuente de datos es todavía limitada.
    - Detección de amenazas: Análisis, alertas y cuadros de mando generados de forma manual.
    - Respuesta ante incidentes: Interfaz limitada en comparación a la siguiente generación.
    - Dashboards y visualizaciones: Conjunto limitado de visualizaciones preelaboradas.

- **3ª Generación**. En esta generación se introducen elementos muy importantes para la detección automática de amenazas y para la reacción proactiva ante diferentes eventos de seguridad complejos.

Estos elementos son los UEBA que analizan comportamientos de forma automática, y los SOAR que son capaces de interactuar de forma automática con los sistemas IT y de seguridad.

    - Escalabilidad: Basada en Data Lake. Escala de forma ilimitada.
    - Retención de datos: Retención histórica ilimitada. Se incluyen nuevas fuentes de origen como las aportadas por la nube.
    - Recolección de datos: Ingesta automática de cualquier fuente de datos.
    - Detección de amenazas: Automatizado. Basado en machine learning y perfiles de comportamiento.
    - Respuesta ante incidentes: Se integran con las herramientas IT y de seguridad. Se posee una capacidad completa de orquestación y automatización (SOAR).
    - Dashboards y visualizaciones: Exploración completa de datos de tipo Business Intelligence (BI).

- La próxima generación de los SIEM.

En esta nueva generación de SIEM, estos se integran con las herramientas IT de seguridad de la organización, completándose de esta forma las capacidades de los SOAR. Podemos por tanto encontrarnos con las siguientes capacidades:

- **Identificación de amenazas complejas**. Esto se basa en los perfiles de comportamiento automáticos. Esto consigue que se puedan detectar comportamientos que sugieran una posible amenaza.

- **Movimientos laterales**. Dado que se analizan todos los datos de la red a través de sus eventos, y a su vez de los múltiples sistemas que componen la organización, los SIEM son capaces de detectar movimientos laterales.

- **Análisis del comportamiento de la entidad**. Todos los activos críticos poseen patrones de comportamiento únicos. Por ello los SIEM pueden aprender estos patrones y descubrir automáticamente anomalías que puedan ser fruto de un incidente.

- **Detección sin reglas ni firmas de amenazas**. Esto se consigue en base al uso del aprendizaje automático que es capaz de detectar incidentes sin una detección de tipo precursora.

- **Respuesta automática ante incidentes**. Una vez detectado un cierto evento de seguridad, se puede ejecutar una secuencia de acciones específicas determinadas con antelación (playbooks) para contener y mitigar el incidente. El SIEM por tanto se están convirtiendo en herramientas SOAR.

2.6. Buenas prácticas.

Vamos a describir una serie de buenas prácticas a la hora de implementar una solución SIEM de forma exitosa. Los siguientes puntos son recomendaciones realizadas por el **Infosec Institute**.

Definición de requisitos:

- Hay que definir los requerimientos para la monitorización, los reportes y las auditorías. Se debe consultar a todos los actores implicados dentro de la organización antes de implementar el SIEM.

- Determinar el alcance del sistema SIEM. Qué partes de la infraestructura va a cubrir, las credenciales necesarias para poder enviar los logs y cuánta información de log se va a enviar.

- Definir qué datos se van a recolectar, cómo de accesibles son, la retención temporal de estos, la forma de asegurar la integridad de los datos almacenados, los datos probatorios de incidentes y la forma de tratar los datos históricos o privados.

Monitorizar y reportar lo siguiente:

- **Monitorización de acceso**. Detección de accesos anómalos a recursos clave de la organización.

- **Defensa perimetral**. El estado de las defensas perimetrales, posibles ataques y cambios de configuración que generen riesgos.

- **Integridad de los recursos**. La integridad de los recursos críticos de red. Sus estados, copias de seguridad, gestión de cambios, amenazas detectadas y vulnerabilidades.

- **Detección de intrusiones**. Los incidentes reportados por los IDS/IPS o generados a partir del análisis o correlación sobre los eventos del SIEM.

- **Defensa contra malware**. Violación, amenazas o actividades relacionadas con los sistemas que detectan amenazas.

- **Defensa de aplicaciones**. Estados, cambios de configuración, violaciones y anomalías de servidores web, base de datos y otros recursos de las aplicaciones web.

- **Uso aceptable**. Estado, problemas y violaciones relacionadas con el uso aceptable, obligatorio o medido de los recursos de la organización.

Controles críticos para implementar dentro del SIEM.

Por último, pasamos a enumerar los 20 controles críticos que se pueden monitorizar a través de un SIEM. Esta lista de controles fue elaborada por el SANS Institute junto al Center of Strategic and International Studies (CSIS).

Estos controles críticos se centran básicamente en cuatro filosofías para combatir las amenazas existentes dentro de los sistemas.

- Las defensas deben enfocarse en abordar los ataques más comunes y dañinos a día de hoy y en un futuro cercano.
- Los entornos empresariales deben garantizar controles consistentes a lo largo de toda la empresa para evitar los ataques.
- Las defensas deben automatizarse en la medida de lo posible y medirse de forma periódica, utilizando técnicas también automatizadas cuando esto sea posible.
- Para poder abordar la problemática de los ataques actuales, las organizaciones deben de implementar una variedad de técnicas específicas que generen una defensa más efectiva.

Pasamos por tanto a describir cada uno de los controles a monitorizar.

- **Control crítico 1. Inventario y control de hardware de los activos**: Gestionar activamente todos los dispositivos hardware de la red (inventario, seguimiento y corrección) para que solo los autorizados tengan acceso. De esta forma se pueden llegar a su vez a detectar dispositivos no autorizados y se les impidan su acceso.

- **Control crítico 2. Inventario y control de activos de software**: Administrar de forma activa todo el software de la empresa, para que solo el autorizado esté instalado y se pueda ejecutar, y aquel que no lo esté sea detectado para evitar su instalación, ejecución o se elimine en caso de estar instalado.

- **Control crítico 3. Gestión continua de vulnerabilidades**: Identificación, remediación y minimización de la ventana de exposición de las distintas vulnerabilidades. Todo ello de forma proactiva.

- **Control crítico 4. Uso controlado de privilegios administrativos**: Esto se consigue a través de los procesos y herramientas utilizadas para trazar/controlar/prevenir/corregir el uso, asignación y configuración de privilegios administrativos en ordenadores, redes y aplicaciones.

- **Control crítico 5. Configuraciones seguras para hardware y software en dispositivos móviles, computadoras portátiles, estaciones de trabajo y servidores**: Se deben de implementar y administrar de forma activa la configuración de seguridad de estos dispositivos a través de la implementación de un proceso de gestión de

la configuración y de los controles de cambios producidos para evitar que los atacantes exploten servicios y configuraciones vulnerables.

- **Control crítico 6. Mantenimiento, monitoreo y análisis de registros de auditoría**: Recopilación, administración y análisis de los registros de auditoría de eventos que pueden ayudar a detectar, comprender o recuperarse de un ataque.

- **Control crítico 7. Protecciones de correo electrónico y navegador web**: Minimización de la superficie de ataque y de las oportunidades de ataques de ingeniería social a través del correo electrónico o las navegaciones web.

- **Control crítico 8. Defensa frente al Malware**: Control de la instalación, difusión y ejecución de código malicioso en múltiples puntos dentro de la organización. Uso de la automatización para permitir las actualizaciones rápidas de las defensas, la recopilación de datos y las acciones correctivas.

- **Control crítico 9. Limitación y control de red**: Se deben controlar continuamente el uso de puertos, protocolos y servicios en los distintos dispositivos de red para minimizar posibles ventanas vulnerables de las que se pueden beneficiar los atacantes.

- **Control crítico 10. Capacidad de recuperación de datos**: Se trata de los procesos y herramientas para respaldar adecuadamente la información crítica usando una metodología segura para el éxito en la recuperación en caso de necesidad.

- **Control crítico 11. Configuraciones seguras para los dispositivos de red**: Se debe de administrar la configuración de seguridad de los dispositivos relacionados con la infraestructura de red (Firewalls, Routers y Switches) mediante un riguroso proceso de gestión de la configuración y control de los cambios para evitar la explotación de servicios y configuraciones vulnerables.

- **Control crítico 12. Defensa del perímetro**: Relacionado con la detección/prevención/corrección relacionado con el flujo de información que se transfiere entre redes de diferentes niveles de confianza con un enfoque en los datos que pueden ser dañinos para la seguridad.

- **Control crítico 13. Protección de datos**: Relacionado con los procesos y herramientas utilizados para evitar la filtración de datos, el mitigar los efectos causados cuando ya se han producido y garantizar la privacidad e integridad de la información confidencial.

- **Control crítico 14. Acceso controlado basado en la necesidad de conocimiento**: Son los procesos y herramientas utilizados para detectar/controlar/prevenir/corregir el acceso seguro a los activos críticos (información, recursos, sistemas) de acuerdo con el principio de qué

personas, sistemas y aplicaciones tienen la necesidad y el derecho para acceder a estos activos críticos, todo ello basado en una clasificación que debe de estar aprobada.

- **Control crítico 15. Control de acceso inalámbrico**: Procesos y herramientas utilizadas para detectar/controlar/prevenir/corregir todo lo relacionado con el uso y seguridad de las redes inalámbricas, puntos de acceso y sistemas cliente.

- **Control crítico 16. Monitorización y control de cuentas**: Administración activa del ciclo de vida de las cuentas de los sistemas y aplicaciones. En concreto, su creación, uso, inactividad y eliminación, para minimizar la superficie de ataque.

- **Control crítico 17. Implementación de un programa de captación y concienciación sobre seguridad**: Orientado a todo el personal de la organización y priorizando aquellos roles críticos, identificar los conocimientos, habilidades y capacidades específicas necesarias para una correcta defensa en cuestión de seguridad dentro del factor humano. Por ello se debe de desarrollar y ejecutar un plan integrado para poder evaluar, identificar brechas y remediarlas a través de políticas, planificación, capacitación y programas de concienciación.

- **Control crítico 18. Seguridad referente al software de los aplicativos**: Administración del ciclo de vida referente a la seguridad de todo el software desarrollado y adquirido internamente por la empresa, para prevenir, detectar y corregir las debilidades a nivel de seguridad.

- **Control crítico 19. Respuesta y gestión ante incidente**s: Proteger la información de la organización, así como su reputación, mediante el desarrollo e implementación de una infraestructura de respuesta ante incidentes de seguridad, en los cuales se detecte rápidamente un posible incidente y se pueda contener su daño de manera efectiva, eliminado la acción del atacante y restaurando la integridad de los sistemas y/o redes afectados.

- **Control crítico 20. Pruebas de penetración y de Red Team**: Se trata de la prueba general de las defensas de la organización mediante la simulación de los objetivos y acciones de un atacante.

En base a estos controles críticos y usando una plataforma SIEM, una organización puede desplegar un control efectivo sobre la eficacia de sus defensas a nivel de seguridad. El objetivo es poder automatizar dentro de las funcionalidades del SIEM los controles anteriores (siendo 15 de estos los más fácilmente automatizables). Se por ello intentar automatizarlo dado que el punto débil normalmente suele ser la acción humana sobre recopilación, interpretación y respuesta ante las amenazas. Sin embargo, con la automatización se consigue tener un mayor éxito en todo este proceso.

2.7. Arquitectura de los SIEM

Los sistemas SIEM pueden ser dividirse a nivel de arquitectura en dos vertientes, la física y la lógica. A continuación, pasaremos a ver en detalle ambas.

2.7.1. Arquitectura lógica.

A nivel lógico, la arquitectura ya ha sido analizada en el anterior punto. Pasamos a enumerar los 12 componentes que lo conforman:

- Recolección y agregación de datos.
- Feeds de tipo Threat Intelligence.
- Correlación y monitorización de eventos de seguridad.
- Analítica.
- Alertas.
- Dashboards.
- Compliance.
- Retención.
- Análisis forense de logs.
- Threat Hunting.
- Respuesta ante incidentes.
- Automatización del SOC.

2.7.2. Arquitectura física.

Una vez vista la parte lógica de la arquitectura de un SIEM, pasamos a describir la arquitectura física de la que se componen los SIEM.

<u>El proceso de administración de los logs.</u>

Un SIEM en su núcleo no es más que una plataforma para gestionar logs. Esta administración implica recopilar los datos, administrarlos para permitir su análisis y su retención a lo largo del tiempo.

- **Recopilación de datos**. Los SIEM recopilan logs y eventos de decenas, sino cientos o miles de sistemas dentro de una organización. Un dispositivo cualquiera genera un evento cada que en él sucede algo, y lo recopila dentro de un archivo de texto plano o en una base de datos. Los SIEM pueden recopilar los datos de cuatro maneras distintas:

    - A través de un agente instalado dentro del dispositivo a monitorizar. Este es quizá el método más común.
    - Conectándose directamente con el dispositivo mediante un protocolo de red o una llamada a la API.

- Accediendo a los logs de registro directamente desde el almacenamiento. Esto se realiza generalmente en formato Syslog.
- A través de un protocolo de transmisión de eventos como puede ser SNMP, Netflow o IPFIX.

Como objetivo, el SIEM debe recopilar los datos de los distintos dispositivos, normalizarlos y guardarlos en un formato que permita su posterior análisis.

- **Gestión de datos**. Dada la enorme gran cantidad de datos que puede llegar a almacenar un SIEM, debemos tratar los datos de la siguiente manera:

  - Se deben almacenar. Ya sea dentro de las propias instalaciones de la empresa, en la nube, o en ambos.
  - Optimizado e indexado. Para poder permitir un correcto análisis y exploración de los datos, los logs deben de ser optimizados y almacenados dentro de índices.
  - Almacenamiento en jerarquía. Los datos importantes para la supervisión en tiempo real deben de estar almacenados dentro de un dispositivo de alto rendimiento, mientras que aquellos que son menos probables de ser consultados (y que son la mayoría), deben de serlo dentro de almacenamientos de gran capacidad, pero no tan eficientes en cuanto a su velocidad. Estos últimos son mucho más económicos.

  Los SIEM actuales cubren la problemática de la necesidad de espacio casi ilimitado a bajo coste con altas prestaciones mediante el uso de tecnologías como la de Data Lake.

- **Retención de los logs**. Existen estándares dentro de la industria como pueden ser PCI DSS, HIPPA y SOX requieren que los registros se almacenen entre 1 y 7 años. Dada la gran diversidad de fuentes que almacena el SIEM, éste debe ser capaz de retener en función de criterios bien definidos los períodos de retención de cada una de las fuentes.

  Por todo ello, los SIEM utilizan estrategias varias para reducir el volumen de los logs almacenados.

  - **Servidores de Syslog**: Syslog es un estándar que normaliza los logs y conserva únicamente la información esencial en un formato estandarizado. Syslog permite comprimir los logs a la vez que retiene grandes cantidades de datos históricos.
  - **Eliminación programada de logs**: Los SIEM eliminan automáticamente los logs antiguos que ya no son necesarios para su propósito. Hay que indicar que la rotación de logs es definitiva. Es su último estado dentro de su ciclo de vida.
  - **Filtrado de logs**: no todos los registros son realmente necesarios para los requisitos de cumplimiento que enfrenta su organización o

para fines forenses. Los registros se pueden filtrar por sistema de origen, tiempos u otras reglas definidas por el administrador de SIEM.

- **Sumarización de logs**: Los logs se pueden reducir en tamaño para mantener solo los elementos más relevantes. Esto puede llegar a reducir significativamente el espacio almacenado.

Una de las grandes ventajas que se está utilizando en referencia a la posibilidad de poder almacenar gran cantidad de logs durante largos períodos de tiempo es, a parte de para su análisis forense, y cumplimiento normativo, para su uso en la analítica del comportamiento a través del machine learning. Esto ha generado la incorporación dentro de los SIEM de módulos de UEBA. Dado que actualmente se pueden almacenar más cantidad y mayor variedad hasta cubrir todo el espectro de eventos de seguridad, se está consiguiendo detectar una gama más amplia de anomalías e incidentes de seguridad.

El flujo (Flow) de los logs.

Un SIEM puede llegar a capturar todos los logs de una organización. Esto puede suponer millones de eventos al día. Pero esto es solo la entrada del sistema, la salida de todo este proceso puede llegar a ser unas pocas (o incluso ninguna) alertas. Por tanto, hay todo un trabajo complejo por debajo desde el mismo momento en que se ingieren dichos logs.

Todo este proceso es un filtrado de ruido para realmente mantener únicamente los datos relevantes. El filtrado se puede realizar en varios puntos del flujo, puede ser un filtro durante la recolección, pero también puede ser filtrado en un momento posterior al indexado. En todo caso se intenta optimizar todo el proceso para obtener datos relevantes que permitan un análisis ágil. Estos eventos filtrados y como se ha comentado relevantes, se correlacionan y analizan con mayor profundidad. De todo este último proceso, únicamente aquellos que superen los umbrales indicados se convertirán en alertas de seguridad.

Un posible escenario del flujo del log podría ser el siguiente:

- Captura.
- Filtrado (puede ir posterior al siguiente punto).
- Indexado.
- Analizado.
- Correlado.
- Umbral de disparo.

En todo el proceso anterior se ha remarcado que el proceso de filtrado es básico. Esto es así dado que el volumen de logs es tan grande que puede llegar a dificultar su análisis. El filtrado en origen es simplemente indicarle al agente recolector qué eventos debe ingerir (también se pueden realizar en

elementos intermedios e incluso justo antes de la indexación). El problema que surge con este método es que o bien este proceso está muy bien definido, o puede surgir la posibilidad de pérdida de eventos relevantes. De hecho, es muy complicado determinar si un evento determinado puede llegar a ser crítico o no en el futuro y en referencia a un posible incidente de seguridad.

Los sistemas SIEM facilitan el filtrado de otra forma. Permiten la indexación de todos los logs que deseemos para posteriormente y en base a una serie de estructuras de datos, facilitar mediante las búsquedas encontrar los datos de forma sencilla y rápida. Un ejemplo sencillo es el concepto de índice. Cada uno de los logs que se almacenan dentro del sistema, lo hacen dentro de unas estructuras de datos definidas en base a unas etiquetas. La principal suele ser el índice al que hace referencia un determinado log. Por ejemplo, todos los eventos de canal seguridad de un sistema Windows podría indexarse dentro de un índice llamado “windows”. De esta forma estamos utilizando la forma en como estructuran los datos internamente los SIEM para poder posteriormente filtrar de forma eficiente los logs almacenados, obteniendo así los resultados deseados.

El ciclo de vida de los logs.

A modo resumen podemos determinar que un log a lo largo de su existencia dentro de la infraestructura SIEM. Hay que entender que un log es un evento generado dentro de un sistema y que se integra bajo unas condiciones dentro del SIEM pasando de este modo a iniciarse un ciclo que le llevará hasta su eliminación.

Los logs por tanto pasan por los siguientes estados:

- **Generación**. Se produce dentro del dispositivo origen.

- **Almacenamiento**. Una vez el log ha llegado desde el dispositivo, se ha almacenado dentro de la estructura de datos de almacenamiento del SIEM. En esta fase es cuando por regla general se producen sobre estos todos los procesos típicos del SIEM como pueden ser el análisis, la monitorización, … Por tanto, el log está en su fase principal en cuanto a utilidad.

- **Archivado**. Los SIEM tienen la capacidad durante la gestión de logs de archivar aquellos logs que cumplan alguna condición como puede ser un tiempo de envejecimiento. Esto se suele realizar por un tema de optimización de recursos. El log no posee la capacidad de ser consultado, pero todavía existe dentro del sistema de almacenaje del SIEM. En caso de necesidad, este log puede volver al estado de Almacenamiento para que sobre éste puedan realizarse las acciones oportunas.

- **Eliminación**. Todos sistema SIEM en su gestión del ciclo de vida posee esta última fase en la que elimina de forma definitiva el log de sus sistema de almacenaje. La eliminación normalmente está planificada y

es fija en base a una serie de criterios como pueden ser el tiempo que llevan almacenados o el volumen que están ocupando. Indicar que un log también puede ser eliminado antes de que se cumplan estos criterios. Una vez eliminado, no puede ser recuperado porque se produce un proceso de rotación en la que se sobrescribe la información eliminada por una nueva.

Fuentes de logs.

Veamos a continuación una serie de fuentes que debemos tener en cuenta a la hora de integrarse dentro de los sistemas SIEM para poder cumplir con sus funcionalidades.

- **Eventos de seguridad**. Se tratan de los eventos relacionados con los dispositivos de seguridad. Estos dispositivos pueden ser, por ejemplo: IDS, EDR, DLP, FWs, …

- **Logs de red**. Aquí se enmarcan los logs relacionados con los dispositivos de red como pueden ser los Routers, Switches, servidores DNS, VPC, …

- **Aplicaciones y dispositivos**. Se deben recoger también logs de aplicaciones y dispositivos como pueden ser bases de datos, aplicaciones web, laptops y desktops de usuario final, dispositivos móviles, …

- **Infraestructura IT**. Otras fuentes interesantes para añadir dentro del SIEM son configuraciones de las infraestructuras, usuarios, inventario de software, …

Los nuevos sistemas SIEM pueden actualmente acceder a los datos contenidos dentro de infraestructuras de la nube (AWS o Azure) o aplicaciones SaaS (SalesForce o Google Apps).

Modelo de despliegue SIEM. MSSP vs In-house.

A la hora de implementar un sistema SIEM a nivel de arquitectura, podemos elegir entre dos modelos distintos de gestión. Estos dos modelos pueden estar separados o mezclados entre sí. Estamos ante un modelo MSSP (managed security service provider) que consiste en que el servicio es prestado por un proveedor de servicios externo a la compañía. El otro modelo es gestionarlo de forma interna. Ambos modelos pueden ser complementarios a la vez como se verá a continuación.

- **Modelo eventos de seguridad**. Este modelo es el tradicionalmente usado. El SIEM se aloja dentro del Data Center de la organización, a menudo siendo un dispositivo dedicado. Los sistemas de almacenamiento y su administración son responsabilidad de la organización. Todos los elementos componentes de un SIEM son gestionados internamente por la empresa. El mantenimiento de este tipo

de infraestructura puede llegar a ser muy compleja y costosa de mantener.

- **Modelo Cloud SIEM, autogestionado**. En este modelo el MSSP se encarga de gestionar la parte referente a la recepción de los eventos del sistema, su recopilación y su agregación. La organización se encarga de la correlación, el análisis, administración de las alertas y Dashboards, y todos los procesos relacionados con los resultados a nivel de seguridad obtenidos.

- **Modelo auto-hospedaje, híbrido-administrado**. El MSSP junto con su personal de seguridad se encargan de la implementación de la recolección y la agregación de los eventos del SIEM, de la correlación, del análisis de los datos, de las alertas y los Dashboards. La empresa es la encargada de la compra de toda la infraestructura del SIEM a nivel de hardware y software. Por tanto, en este modelo la parte principal del servicio recae sobre el MSSP.

- **Modelo SIEM as a Service**. En este último modelo el MSSP se encarga de la recopilación de los eventos, de la agregación, correlación, análisis, de las alertas y los Dashboards. La organización de forma interna se encarga de gestionar todos los procesos de seguridad que se aprovechan de los datos suministrados por el SIEM. Es decir, la parte de la inteligencia aplicada a la organización.

Basándonos en estos modelos, cada organización debe decidir cuál de estos es el que más le conviene implementar. Ya sea por una cuestión económica, de recursos (existencia previa de una infraestructura SIEM dentro de la organización, existencia de personal especializada para gestionarla o si se pueden enviar los datos fuera de la organización) o ambas. El MSSP lo que proporciona por tanto es un servicio profesional especializado para la gestión de un sistema complejo.

Como se ha podido ver en los párrafos anteriores, una solución SIEM puede estar creada tanto On-premise, como en la nube como en un modelo mixto. Dependiendo de las necesidades de cada organización, así se deberá adoptar un modelo u otro.

Dimensionamiento del SIEM.

Dada la gran cantidad de logs que recogen los sistemas SIEM hoy en día, y que generalmente se almacenan dentro de las organizaciones, se requiere que las organizaciones consideren cuidadosamente el tamaño de los logs y eventos que se están generando, y los recursos a nivel de sistemas necesarios para administrar todo esto.

- EPS (Events Per Second).

Se introduce aquí el concepto de EPS (Events Per Second). Esto nos servirá para calcular la velocidad con la que se generan los logs. El EPS se calcula de la siguiente forma:

EPS = # eventos seguridad / período de tiempo en segundos

Hay que entender que los EPS se mueven entre tiempos valle y una serie de picos. Un ejemplo sencillo puede ser los eventos de un FW perimetral. Cuando se produce un ataque puede generar durante cierto tiempo un pico importante en el número de eventos generados por el dispositivo.

Por lo anterior, las organizaciones deben lograr un equilibrio entre las mediciones normales y máximas generadas por cada uno de los dispositivos. Esto es muy importante dado que a la hora de diseñar el SIEM a nivel de almacenamiento deberemos tener en cuenta estos factores.

- Modelo simple de predicción del valor EPS durante tiempos normales y picos.

  A continuación, vamos a ver cómo podemos calcular un valor predictivo del EPS a lo largo del tiempo contando con los valores valle y pico. Para ello vamos a seguir los siguientes pasos:

  - Medir el EPS normal y máximo. Esta observación se realizará durante 90 días sobre el sistema objeto.
  - Se debe de estimar la cantidad de picos por día.
  - Se estima la duración en segundos de dichos picos, y por extensión, el total de segundos por pico y día.
  - Calculamos el total de eventos pico por día mediante la siguiente fórmula: Eventos pico por día = (Total segundos pico por día) * EPS del pico.
  - Calculamos el total de eventos normales por día siguiendo la fórmula siguiente: # total de eventos normales por día = (Total de segundos - Total de segundos pico por día) * EPS normal.
  - La suma de estos dos números es la velocidad total estimada.
  - Se recomienda a su vez a este valor añadirle un 10% para el Headroom, y otro 10% para un crecimiento.
  - Por tanto, el número final de eventos por día se podrá expresar de la siguiente manera:
  - (Total eventos pico x día + total eventos normales x día) * 110% de Headroom * 110% de crecimiento.

Por todo ello, para dimensionar un SIEM lo que se debe de realizar es un inventario de los dispositivos de los que se tiene la intención de recopilar logs. Una vez realizado lo anterior, se debe de multiplicar el número de dispositivos similares por su EPS estimado. De esta forma se obtiene el número total de logs por día.

Existe a nivel de necesidad de almacenamiento una regla general que dice que un log promedio ocupa unos 300 bytes. Por ello, por cada 1.000 EPS (86,4 millones de eventos por día), el SIEM necesita almacenar unos 26 GB por día, 777 GB al mes, 9.331 GB al año.

Son como se puede apreciar, capacidades muy grandes.

Dimensionamiento de hardware.

Una vez determinada la velocidad y el volumen de los eventos, hay que considerar los siguientes factores para dimensionar correctamente el hardware de un SIEM.

- **Formato de almacenamiento**: Se debe determinar cómo se van a almacenar los logs dentro del sistema. Se pueden almacenar dentro de un archivo plano, una base de datos relacional o un almacén de datos no estructurado como puede ser Hadoop.

- **Implementación de almacenamiento y hardware**: Se debe conocer si se pueden mover los datos a la nube. Muchas empresas por regulación o por políticas internas no pueden exponer sus datos a un almacenaje fuera de sus instalaciones. El almacenaje de los logs en la nube es una opción a nivel de gestión muy atractiva para las empresas. En caso de no poder o no querer decantarse por esta opción, la mejor opción es almacenar los datos dentro de sistemas como Hadoop, las bases de datos no relacionales o los dispositivos de almacenamiento de alto rendimiento.

- **Compresión de los datos**: Debemos conocer el factor de compresión usado por los SIEM. Los logs no se almacenan de sin más, sino que se comprimen para optimizar su almacenaje.

- **Cifrado**: Se debe determinar la necesidad de cifrar los datos. Esto puede ocurrir en situaciones en las que los datos requieran una mayor seguridad por cumplimiento normativo o de políticas.

- **Requisitos de software y hardware**: Hay que tener en cuenta que los SIEM poseen entre otras, dos funcionalidades bien diferenciadas a la hora de procesar los datos. Por un lado, tienen la posibilidad de realizar consultas y monitorización en tiempo real, ello obliga a que los datos más recientes estén disponibles de forma muy rápida. Por ello, en su almacenamiento se debe de contemplar esta necesidad. Y, por otro lado, también poseen la funcionalidad de retención de datos durante largos períodos de tiempo. Esto hará que surja la necesidad de un almacenaje de alto volumen y bajo coste. Se puede apreciar claramente en los SIEM actuales cómo cuando se realizan consultas de datos recientes, los resultados son obtenidos con mayor velocidad que si se consultan datos más antiguos y que están almacenados posiblemente en sistemas más lentos (dependerá obviamente en cómo se haya implementado a nivel hardware su arquitectura).

- **Failover y backup**: Al tratarse de un sistema crítico para la organización, el SIEM debe ser diseñado con alta disponibilidad, teniendo a su vez un sistema bien planificado de copias y respaldos para posibilitar la continuidad del negocio.

Escalabilidad y Data Lakes.

Como se ha visto hasta el momento, los SIEM almacenan una cantidad ingente de logs. Dada la gran cantidad de datos, y la necesidad de recuperarlos con velocidad, se han adoptado una serie de soluciones para solucionar este problema. El Data Lake no sustituye en ningún momento al SIEM dado que siempre serán necesarias sus grandes capacidades para analizar los datos en el ámbito de la seguridad. Por ello, el Data Lake complementa al SIEM en los siguientes aspectos:

- Almacenamiento casi ilimitado y de bajo coste basado en dispositivos de tipo commodity.
- Nuevas formas de procesar grandes datos: las herramientas pertenecientes a Hadoop (como son Hive y Spark), permiten el procesamiento rápido de grandes cantidades de datos, al tiempo que permiten que la infraestructura tradicional de SIEM consulte los datos a través de SQL.
- La posibilidad de retener todos los datos necesarios de una gran multitud de nuevas fuentes de datos, como son las aplicaciones en la nube, IoT y los dispositivos móviles.

Hoy día existen otras alternativas a Hadoop como son ElasticSearch, Cassandra y MongoDB.

Otra gran ventaja de este tipo de almacenamiento es que los costes a nivel de hardware se reducen mucho. Para crecer, simplemente se añaden nuevos nodos que son ejecutados en hardware básico o alojado en la nube. Es un crecimiento de tipo horizontal.

Por tanto, los SIEM que se basan en este modelo tienen la ventaja de poder agregar nuevas fuentes de forma sencilla o incrementar el período de retención de logs a bajo coste.

Informes SIEM. Dashboards y visualización.

Uno de los objetivos principales de un SIEM es la generación de información útil y procesable destinada para los equipos de seguridad. Esta información puede mostrarse a través de los siguientes elementos:

- **Paneles de control**: Muestran el estado de los sistemas y las métricas relacionadas con la seguridad, resaltando los posibles problemas de seguridad.

- **Alertas y notificaciones**: Sirven para que los equipos de seguridad puedan investigar una anomalía o problemática de seguridad.

- **Las API y los servicios web** permiten el uso de sistemas externos, como el BI y análisis de comportamiento herramientas, para acceder a los datos de SIEM y analizarlos desde nuevas perspectivas.

- **Exploración y análisis de datos**: Se trata de la exploración proactiva de los eventos de seguridad en busca de amenazas (Threat Hunting) o para investigar un incidente de seguridad.

- **Machine learning**: Los SIEM actuales utilizan técnicas de machine learning para ayudar a los equipos de seguridad a recopilar datos relevantes en base a anomalías detectadas y que pueden ser relevantes para la detección o el análisis de un posible incidente de seguridad.

Arquitectura SIEM: entonces y ahora.

Históricamente, los SIEM han sido una infraestructura demasiado costosa y monolítica para la empresa, basada en software propietario y hardware personalizado para poder manejar los grandes volúmenes de datos.

El objetivo a lo largo de los años ha sido el que los sistemas SIEM sean más ágiles y livianos, a la par que más “inteligentes”.

Por ello, las nuevas arquitecturas son más asequibles, más fáciles de implementar y presta a los equipos de seguridad un gran ayuda a detectar posibles incidentes de seguridad más rápidamente o incluso a detectar nuevas amenazas que de otra forma no podrían haber sido encontradas.

La nueva generación destaca por los siguientes atributos:

- **Uso de tecnología moderna como es el Data Lake**: Esta solución ofrece un gran almacenamiento de datos con una escalabilidad casi ilimitada a bajo coste y un rendimiento mejorado.

- **Nuevas opciones de hosting y opciones de administración**: Los MSSP ayudan a las organizaciones a implementar los SIEM, encargándose de parte de la infraestructura (de forma local o en la nube) y prestando su servicio profesional en base a su expertise administrando estos sistemas.

- **Escalabilidad dinámica y costes predecibles**: Se simplifica todo lo relacionado con la complejidad de los cálculos a nivel de arquitectura, tanto en los inicios como cuando se debe crecer. El almacenamiento de los sistemas SIEM crecen de forma dinámica y previsible.

- **Nuevas ideas como los UEBA**: Las nuevas arquitecturas SIEM incluyen componentes de análisis avanzados que van más allá de las correlaciones tradicionales para descubrir en base a la gran cantidad de

datos almacenados y estos tipos de algoritmos, nuevas anomalías y posibles incidentes de seguridad.

- **Potenciar la respuesta ante incidentes**: los SIEM modernos aprovechan la tecnología de Orquestación y Automatización de Seguridad (SOAR) para identificar y responder automáticamente ante incidentes de seguridad. Además, ayuda al personal del SOC en la investigación de incidentes de seguridad.

2.8. Agregación, procesamiento y análisis de logs de seguridad.

Los logs y los eventos de seguridad son la base para la supervisión, la investigación y el análisis forense a nivel de seguridad.

Veamos en profundidad todo este proceso empezando por la agregación de logs.

Agregación de logs.

La agregación de logs es el proceso por el cual se recopilan múltiples logs de diferentes sistemas, son analizados y se extraen sus datos de manera estructurada. Además, son almacenados en un formato que las herramientas modernas puedan realizar búsquedas fácilmente sobre estos.

Podemos definir cuatro formas de agregación de logs habituales. Pudiéndose a su vez darse combinaciones de estos.

- **Syslog**. Se trata de un protocolo estándar de logging. Se puede configurar un servidor Syslog para que reciba los logs de múltiples sistemas, almacenándolos en un formato eficiente y condensado para sean fácilmente consultables. Los agregadores de logs pueden leer y procesar directamente los datos de tipo Syslog.

- **Streaming de eventos**. Nos estamos refiriendo a los protocolos como SNMP, Netflow e IPFIX. Dichos protocolos permiten a los dispositivos de red transmitir información de manera estándar referentes a sus actividades. El agregador puede en estos casos interceptar, analizar y agregar toda esta información sobre los logs dentro del almacenamiento central.

- **Recolectores de logs**. Son una serie de agentes de tipo software que se ejecutan en los dispositivos, capturan toda la información, la analizan y las envían al agregador centralizado para su almacenamiento y análisis.

- **Acceso directo**. Los agregadores de logs pueden acceder de forma directa a los dispositivos a monitorizar a través de una API o un protocolo de red para recibir de forma directa dichos logs. Este tipo de recolección requiere de una integración distinta por cada una de las fuentes de datos.

Procesamiento de logs.

El procesamiento consiste en recopilar los logs de múltiples fuentes sin procesar, identificar su estructura o esquema y transformarlos en una fuente de datos coherente y estandarizada.

El flujo de procesamiento de logs es el siguiente:

- **Parseo de logs**. Cada log tiene un formato de datos específico que pueden repetirse según la fuente y tipo de la que provenga, y que incluyen campos y valores de datos.

  Un parseador es por tanto un componente software que tiene como entrada un log determinado de una fuente y lo convierte en datos estructurados. Este software es capaz de parsear múltiples tipos de fuentes de sistemas comunes.

- **Normalización y categorización de logs**. La normalización consiste en la conversión de cada dato contenido dentro de un log a una representación y categorización consistente a todo el sistema de logs administrados. La mayoría de los logs poseen información básica en común. De esta forma lo que se consigue es poder relacionar eventos entre sí procedentes de distintas fuentes.

  La categorización consiste en dar una diferenciación característica a los eventos. De esta forma se identifican logs relacionados entre sí basados en una serie de etiquetas. Por tanto, se pueden llegar a agrupar conjuntos de logs por categorías.

- **Enriquecimiento de logs**. Esta parte hace referencia al enriquecimiento de logs añadiéndole información importante para darle más valor. Un ejemplo claro es si poseemos una IP, mediante un servicio de geolocalización determinar la ubicación para agregar más información dentro del log.

- **Indexación de logs**. Dado el enorme volumen de logs que se tienen que manejar, y la necesidad de realizar búsquedas eficientes, es necesario crear índices en base a atributos comunes para poder realizar lo anterior.

  Esto hace que las consultas de datos puedan ser de un orden de magnitud más rápidas que si tuvieran que analizar todos los logs almacenados.

- **Almacenamiento de logs**. Dado el grandísimo volumen de datos que se están almacenando hoy en día, la forma de almacenarlos está cambiando. Originalmente se almacenaban en un repositorio centralizado, para pasar actualmente a realizarlo a través de soluciones tipo Data Lake.

Los Data Lakes admiten grandes volúmenes de almacenamiento de logs bajo un bajo coste, con crecimiento horizontal. Para poder acceder de forma rápida a los datos que se encuentran distribuidos, se suelen usar motores de procesamiento distribuidos como puede ser MapReduce u otras herramientas modernas de análisis de alto rendimiento.

Tipos de logs.

Casi todos los sistemas generan algún tipo de log. A continuación, vamos a ver alguna de las fuentes más comunes.

- **Endpoint logs**. En estos tipos de dispositivos podemos encontrar múltiples tipos de logs. Podemos encontrarnos logs a nivel de software dentro toda la pila. Desde el hardware, middleware, el sistema operativo, los aplicativos finales, …

- **Logs de dispositivos de red**. Aquí podemos encontrarnos logs provenientes de los Routers, Switches, Firewalls o balanceadores de carga. Sus logs nos proporcionan datos críticos sobre los flujos de tráfico, los orígenes y destinos de las comunicaciones, sus volumetrías, los protocolos usados, … Suelen transmitir estos logs en formato Syslog, pudiéndose capturar y analizar a través de servidores Syslog distribuidos por la red.

- **Eventos de aplicación**. Son los logs de las aplicaciones que se ejecutan dentro de los servidores o dispositivos de usuario final. En los sistemas Windows se almacenan dentro de un almacén de eventos centralizado. Para el caso de Linux, los podemos encontrar dentro de la ruta /var/log. Aquí también entrarían todos aquellos logs de aplicaciones empresariales, servidores de correo, servidores web o bases de datos.

- **Logs IoT**. Se tratan de fuentes recientes que son aquellos dispositivos conectados al Internet de las cosas. Pueden registrar su propia actividad y/o datos de sensores dentro del dispositivo. Este tipo de dispositivos se están convirtiendo en un desafío a la hora de monitorizarlos dado que su visibilidad dentro de la empresa no siempre está demasiado clara. Una solución que se está dando es el almacenamiento de estos logs de forma centralizada en la nube, para ello se está utilizando el protocolo de recopilación denominado syslog-ng que se caracteriza por su portabilidad y la recopilación centralizada de logs.

- **Proxy logs**. Se tratan de los logs referentes al tráfico de los usuarios internos, ya sea a nivel de red local o solicitudes realizadas a través de Internet. Pueden ser por tanto logs generados por los usuarios, como los generados por aplicaciones o servicios. Es importante destacar que para que se cumpla plenamente el objetivo de monitorización, estos dispositivos deben ser capaces de desencriptar e interpretar todo el tráfico cifrado.

Formatos de logs más comunes.

Veamos a continuación los formatos usados más comunes:

- **Formato de logs CSV**. El log es enviado en formato CSV. Es un formato abierto para representar datos en forma de tabla. Las columnas se separan por comas y las filas por saltos de línea.

- **Formato de registro JSON**. Es un formato de texto sencillo que consta de un subconjunto de la notación literal de objetos de JavaScript. Este tipo de formato es más sencillo que por ejemplo el formato XML al poderse escribir una analizador sintáctico (parser) más sencillo.

- **Formato de evento común (CEF)**. Es un estándar abierto de administración de logs que facilita el intercambio de datos relacionados con la seguridad desde múltiples dispositivos. Proporciona a su vez un formato de registro de eventos común, lo que hace que la recopilación y agregación sea más sencilla. CEF usa el formato de mensaje Syslog.

Monitorización de logs.

Dada la gran cantidad contenida dentro de los distintos logs que se almacenan, mediante la monitorización podemos identificar patrones y problemas dentro de los sistemas. El monitorizar significa el analizar los registros, buscar patrones, reglas o comportamientos que nos avisen sobre eventos importantes y que por ello puedan activarse algún tipo de alerta que genere una acción determinada.

Monitorizar puede por tanto ayudar a identificar problemas en sus etapas iniciales. Puede ayudar a descubrir comportamientos sospechosos que podrían ser síntoma de un posible incidente de seguridad. Puede ayudar al registro del comportamiento base de los dispositivos, sistemas y usuarios, con el objetivo de identificar anomalías que requieran de un posterior análisis o investigación.

Registro de eventos de seguridad.

La agregación de logs y su monitorización es una actividad central para los equipos de seguridad de una empresa. Al recopilar y analizar logs provenientes de sistemas críticos y herramientas de seguridad hace que se puedan identificar comportamientos anómalos o sospechosos, pudiendo ser alguno de ellos un incidente de seguridad.

Debemos por tanto tener claro dos conceptos distintos, qué es un evento y qué es un incidente. Un eventos es algo que ha pasado dentro de algún dispositivo. Uno o más eventos pueden llegar a catalogarse como un incidente si se detecta que se trata de un ataque, una violación de las políticas de seguridad, un acceso no autorizado, …

Los eventos comúnmente más relevantes para la seguridad suelen ser los siguientes:

- Reporte por parte de un antivirus en referencia a la detección de un malware dentro de un dispositivo.
- Reporte desde un firewall sobre tráfico prohibido ya sea de entrada o salida.
- Intento de acceso a un sistemas crítico desde una máquina o IP desconocida.
- Intentos fallidos de acceso a un sistema crítico.
- Cambios en los privilegios de un usuario.
- El uso de protocolos y puertos inseguros o prohibidos.

Los incidentes de seguridad más comunes pueden ser los siguientes:

- Correos maliciosos recibidos por la organización y que el usuario interactúa con él.
- Acceso a servidores web maliciosos por parte de los usuarios de la organización.
- Acciones indebida o prohibidas realizadas por un usuario autorizado.
- Accesos no autorizados a recursos o sistemas.
- Intentos de compromiso, denegación de acceso o eliminación de sistemas dentro de la organización.
- Pérdida o robo de equipos pertenecientes a la organización.
- Fuga de datos o infección por malware a través de medios extraíbles.

Todo lo visto anteriormente se consigue mediante el uso de sistemas SIEM. Ya hemos visto como el SIEM se encarga por un lado de la gestión de los logs, y por otro de todo lo relacionado con el análisis, la detección y notificación de posibles incidentes de seguridad.

Los SIEM analizan por tanto todos los datos y establecen relaciones que ayudan a identificar anomalías, vulnerabilidades e incidentes de seguridad. El SIEM tiene un enfoque eminentemente relacionado con la gestión de la seguridad. El SIEM identifica qué eventos pueden llegar a ser relevantes dentro de la seguridad, para posteriormente ser revisados por una analista humano. También proporcionan una serie de visualizaciones a través de paneles que facilitan dicha labor.

Históricamente, un SIEM tradicional se basa en dos técnicas para generar las alertas: Las reglas de correlación, en las que se especifican una serie de secuencias de eventos que si ocurren determinan que estamos ante una anomalía a analizar; y por otro la do la evaluación de vulnerabilidades o riesgos, que lo que implica es escanear la red para detectar patrones de ataque y vulnerabilidades conocidas. El problema de estas técnicas es que generan gran número de falsos positivos y fallan en el objetivo de detectar nuevos tipos de eventos anómalos e inesperados.

Con los SIEM de nueva generación las cosas se mejoran. Los SIEM avanzados utilizan una tecnología denominada UEBA que se aprovecha del aprendizaje automático para observar patrones de comportamiento, establecer líneas base e identificar de manera inteligente comportamientos sospechosos o anómalos.

Todo esto puede ayudar a detectar riesgos que actualmente son desconocidos o de difícil definición a través de las reglas correladas tradicionales. Detecta por tanto con más eficiencia amenazas internas, ataques dirigidos, fraudes y anomalías durante largos períodos de tiempo o en múltiples sistemas de la organización.

Como resumen, indicar que la gestión de logs de manera eficiente siempre ha sido compleja, y dado que cada vez se almacenan más cantidad y variedad de estos, la complejidad aumenta. Por ello los SIEM de próxima generación lo que pueden hacer es ayudar a administrar y extraer de forma más eficiente los logs relevantes relacionados con la seguridad. Como se ha visto, estos SIEM se basan en los Data Lake, los UEBA y capacidades avanzadas de exploración de datos que ayudan a los analistas con el desempeño del Threat Hunting.

2.9. User and Entity Behavior Analytics (UEBA).

Los UEBA son una nueva solución a nivel de seguridad que utiliza tecnologías de analíticas de datos como pueden ser el machine learning y el Deep learning. Estas herramientas son capaces de detectar comportamientos anormales y posibles riesgos por parte de actividades de usuarios, máquinas y otros elementos pertenecientes a la red corporativa.

Las soluciones de este tipo crean perfiles que modelan el comportamiento estándar de usuarios o entidades de un entorno IT. Usando una variedad de técnicas de analítica, este tecnología puede identificar actividades anómalas en comparación con las líneas base establecidas. De este modo, son capaces de detectar amenazas o posibles incidentes de seguridad.

Los UEBA se definen en tres dimensiones:

- **Casos de uso**. Los UEBA proporcionan información sobre el comportamiento de los usuarios y entidades de la corporación. Se encargan de monitorizar, detectar y alertar de las anomalías encontradas. Estas acciones son aplicables a múltiples casos de uso.

- **Fuentes de datos**. Estas soluciones ingieren los datos a analizar desde un repositorio general, como puede ser un Data Lake o un almacén de datos, o a través de un SIEM. Por tanto, no se deben desplegar agentes para este tipo de soluciones.

- **Análisis**. Utilizan una gran variedad de enfoques analíticos: modelos estadísticos, aprendizaje automático, reglas, firmas de amenazas, …

Integración de los UEBA con los SIEM.

Hay una estrecha relación entre ambos sistemas dado que los UEBA se basan en los datos de seguridad para realizar sus análisis. Dichos datos generalmente son recopilados y almacenados en un SIEM.

Según la visión de Gartner de una solución SIEM de próxima generación, un SIEM debe de incluir la funcionalidad UEBA dado que con ésta cubre parte de esas necesidades:

- **Monitorización de usuarios**. Mediante la inclusión de líneas base y análisis avanzado se pueden establecer el escenario base de la actividad de los usuarios y sistemas, consiguiendo de esta forma detectar comportamientos sospechosos.

- **Análisis avanzado**. Uso mediante la aplicación de modelos estadísticos y cuantitativos, como pueden ser el aprendizaje automático y profundo para la detección de actividad anómala. Esta analítica se complementa con las reglas tradicionales y la analítica basada en correlaciones.

Casos de uso de UEBA.

Los UEBA deben de ser capaces de detectar una serie de casos de uso. Se describe a continuación alguno de estos:

- **Detección de Insiders maliciosos**. Esto se realiza en base al comportamiento estándar de un usuario, detectando actividades anómalas de usuarios en base a lo anterior.

- **Cuenta privilegiada comprometida**. Se detecta el comportamiento anómalo de cuentas o máquinas privilegiadas cuando éstas han sido comprometidas, y desde éstas continúan con los ataques contra otros objetivos.

  La tecnología UEBA es capaz de detectar ataques que tradicionalmente son difícilmente detectables. Nos estamos refiriendo a ataques de tipo Zero-day, movimientos laterales, … Esto es así, dado que estos activos para poder realizar sus actividades maliciosas se comportan de una manera distinta a las línea base establecidas.

- **Priorización de incidentes**. Las soluciones UEBA ayudan a los equipos de seguridad a priorizar aquellos incidentes que son particularmente anómalos, sospechosos o potencialmente peligrosos. Los UEBA pueden ir más allá de la generación de líneas base y modelos de amenazas para agregar datos sobre la estructura de la organización. Un ejemplo sería el indicar la criticidad de los activos, roles y los niveles de acceso de cada uno de los componentes de la organización. En base a esto, una pequeña desviación dentro de un grupo crítico valdría para que se

investigase. Para el resto de los casos la desviación debería ser mucho mayor.

- **Prevención de pérdida de datos (DLP) y prevención de fuga de datos**. Las herramientas tradicionales de DLP pueden llegar a generar un gran número de alertas lo que dificulta su normal tratamiento. Las soluciones UEBA pueden tomar dichas alertas, priorizarlas y consolidarlas para determinar qué eventos representan realmente un comportamiento anómalo en comparación con las líneas base conocidas.

- **Análisis de IoT**. Los dispositivos IoT pueden generar grandes problemas para la organización a la hora de monitorizarlos y detectar riesgos. Los UEBA pueden monitorizar un número ilimitado de dispositivos conectados, establecer una línea base de comportamiento para cada uno de estos dispositivos o para un grupo de ellos similares, y detectar si un dispositivo se comporta fuera de lo habitual.

Métodos de análisis de los UEBA.

La principal característica de los UEBA es el uso de la analítica avanzada. Esto quiere decir que involucra varias tecnologías modernas que pueden ayudar a identificar los comportamientos anómalos (incluso en ausencia de patrones conocidos).

- **Aprendizaje automático supervisado**. Se incorpora al sistema conjuntos de comportamientos buenos y malos conocidos. La herramienta aprende a analizar nuevos comportamientos y determinar si es similar al comportamiento bueno o malo.

- **Redes bayesianas**. Combinan aprendizaje automático supervisado con reglas para crear perfiles de comportamiento.

- **Aprendizaje no supervisado**. El sistema aprende el comportamiento normal y detecta el comportamiento que se sale de éste. Este sistema no permite determinar si el comportamiento anormal es bueno o malo, solo que se desvía de lo normal.

- **Aprendizaje profundo**. Permite la selección e investigación de alertas virtuales. El sistema se entrena mediante conjuntos de datos provenientes de alertas de seguridad y su clasificación (buenas o malas), con esta información es capaz de predecir resultados de clasificación para nuevos conjuntos de alertas de seguridad.

- **Aprendizaje automático reforzado/semi-supervisado**. Se trata de un modelo híbrido donde la base es el aprendizaje no supervisado, y las resoluciones de las alertas reales se retroalimentan dentro del sistema para permitir un ajuste fino del modelo y de este modo reducir el ratio de la relación señal/ruido.

Las técnicas anteriores son de tipo heurístico (a diferencia de los anteriores que eran deterministas). Calculan una puntuación de riesgo que no es más que una probabilidad de que un evento represente una anomalía o un incidente de seguridad. Cuando dicho puntaje excede un cierto umbral, el sistema genera una alerta.

Por tanto, podemos decir que los UEBA funcionan en base al análisis holístico a través de múltiples fuentes de datos. De hecho, analiza todos los datos posibles de un usuario o una entidad específica. Un UEBA debe analizar tantas fuentes de datos como sea posible. Esto hará que sea capaz de analizar datos tan dispares de un usuario dado como puede ser el inicio de sesión de este a través del AD, la importancia dentro de la organización del dispositivo en el que se ha logado, la sensibilidad de los archivos a los que ha accedido y su actividad normal o inusual de red o de tipo malware que nos puede llevar a la conclusión de que ha podido haber un compromiso de cuenta de usuario.

Bases de comportamiento y puntajes de riesgo.

Los dos anteriores son dos conceptos fundamentales dentro de los sistemas UEBA. Estos aprenden el comportamiento normal para identificar el comportamiento anormal. Examinan un amplio conjunto de datos para determinar la línea de base o el perfil de comportamiento de un usuario dado.

Cuando se produce una desviación de la línea base, el sistema suma al puntaje de riesgo del usuario o máquina. Cuanto más inusual, mayor es esta puntuación de riesgo. A medida que se van acumulando más y más comportamientos sospechosos, la puntuación se va acumulando hasta que alcanza un umbral, ocurrido esto se escala el posible incidente a un analista para su investigación.

Este enfoque analítico tiene varias ventajas:

- **Agregación**. La puntuación de riesgo se compone en base a múltiples eventos, de esta forma el analista no tiene la necesidad de analizar manualmente gran cantidad de alertas individuales y sin agregar.

- **Reducción de falsos positivos**. Un evento que se desvía ligeramente de la normal por si solo no generará una alerta de seguridad. El sistema requiere múltiples signos de comportamiento anómalo para que se dispare una alerta. Esto hace que se reduzca la cantidad de falsos positivos a la par que ahorra tiempo a los analistas.

Sin duda una de las grandes ventajas de las reglas generadas por los UEBA en contraposición a las reglas correladas contextuales y tradicionales es que es capaz de establecer una línea de base sensible al contexto para cada grupo de usuarios o sistemas a monitorizar.

Análisis de la línea de tiempo e interrelación entre eventos.

Al analizar un incidente de seguridad, la línea de tiempo es un concepto crítico que puede unir actividades aparentemente no relacionadas. No hay que olvidar que hoy en día muchos de los ataques que se producen son procesos, no eventos aislados.

Los UEBA unen datos de diferentes sistemas y flujos de eventos para construir una línea de tiempo completa de un incidente de seguridad. Una vez que el UEBA une todos los datos relevantes dentro de un mismo contexto, procede a asignar puntajes de riesgo de dichas actividades. Esto hace que se detecte el incidente al poderse disparar la alerta en base a la superación del umbral determinado para ello. Sin esa agregación temporal y contextual referente a un actor determinado, se dejarían de detectar múltiples incidentes de seguridad.

2.10. Análisis de la seguridad a través del Big Data.

En el pasado se debía definir las reglas de correlación de forma manual, esto hacía que no fuese el mejor escenario para una correcta gestión de la seguridad.

Todo esto ha cambiado gracias al machine learning. Con esta nueva tecnología, los analistas de seguridad pueden identificar patrones y amenazas que antes no se podía. Sin embargo, para que esta solución sea eficaz, debe de poder analizar una gran cantidad de datos. El gran desafío por tanto es conseguir almacenar todos estos datos de una manera eficiente que permita toda la analítica.

Data Science, Machine Learning y ciberseguridad.

La Data Science o ciencia de los Datos es una nueva disciplina que aprovecha el análisis científico y matemático de los conjuntos de datos, así como la comprensión y exploración humana, con el objetivo de sacar de los datos una deriva comercial.

En el contexto de la seguridad de la información, esta ciencia ayuda a los analistas de seguridad en base a herramientas para hacer un mejor uso de los datos relacionados con la seguridad, al descubrimiento de patrones ocultos y en fin, a comprender mejor el comportamiento de los sistemas que securizan.

El Machine Learning (parte del campo de la IA) dentro del ámbito de la ciberseguridad se encarga de utilizar técnicas estadísticas para permitir que las máquinas aprendan sin ser programadas explícitamente. Por tanto, va más allá de las reglas de correlación antiguas, para examinar patrones desconocidos y usar algoritmos para la predicción, clasificación y generación de información.

Aprendizaje supervisado vs no supervisado.

- **Aprendizaje supervisado**. En este tipo de aprendizaje la máquina aprende de un conjunto de datos que contienen entradas y salidas

conocidas. Se construye un modelo que permite predecir cuáles serán las variables de salida para salidas nuevas y desconocidas.
Las herramientas aprenden a analizar nuevos comportamientos y determinar si los resultados son similares a los comportamientos conocidos anteriormente.

- **Aprendizaje no supervisado**. En este aprendizaje el sistema aprende de un conjunto de datos que contienen solo variables de entrada. Por ello no existe una respuesta a la entrada correcta o incorrecta, sino que es el algoritmo el que tiene que descubrir nuevos patrones en los datos.

Aprendizaje profundo (Deep Learning).

Este tipo de aprendizaje simulan al cerebro humano creando redes neuronales digitales para procesar pequeños datos, para posteriormente obtener una imagen más amplia. Este aprendizaje se aplica comúnmente a datos no estructurados, pudiendo aprender de forma automática las características más significativas de los datos. La mayoría de las aplicaciones que usan este aprendizaje, utilizan el aprendizaje supervisado.

Dentro de la ciberseguridad, el Deep Learning se usa para el análisis del flujo de paquetes y el análisis binario de malware. El objetivo, descubrir patrones de tráfico y programas y a través de esto, identificar posibles actividades maliciosas.

Data Mining (minería de datos).

Se basa en el uso de técnicas analíticas, principalmente Deep Learning, para descubrir información oculta en grandes volúmenes de datos.

Las herramientas de seguridad utilizan estas técnicas para realizar tareas como la detección de anomalías en conjuntos de datos muy grandes, la clasificación de incidentes de seguridad o de los eventos de red, y la posibilidad de predicciones de futuros ataques basados en datos históricos.

Por todo lo visto anteriormente, los SIEM actuales necesitan hacer uso de la analítica de Big Data. Los SIEM actuales recogen una gran cantidad de logs de múltiples fuentes, esto hace que las técnicas de analítica descritas anteriormente sean muy convenientes para ayudar en las labores de la seguridad de las organizaciones. Sin este tipo de tecnologías, los UEBA no pueden realizar sus funciones de forma correcta.

Gracias por tanto al Big Data y a las técnicas como el machine learning, Deep learning, … podemos obtener ventajas tan interesantes como los siguientes:

- **Identificación de amenazas complejas**. El análisis de los datos avanzados puede ver los eventos dentro de un contexto temporal y relacional mucho más avanzado. Esto hace que se detecten actividades sospechosas.

- **Análisis del comportamiento de entidades**. Aprenden el comportamiento normal de los activos críticos, para descubrir a partir de esto anomalías de forma automática.

- **Detecciones de movimientos laterales**. Los SIEM al analizar todos los datos de la red, junto a los recursos de los sistemas que componen la organización, gracias al aprendizaje automático se pueden detectar este tipo de ataques.

- **Amenazas internas**. Se basa en la identificación de comportamientos anormales de usuarios o sistemas, conectando entre sí eventos aparentemente no relacionados. De esta forma se detectan ataques más complejos.

- **Detección de nuevos tipos de ataques**. El análisis avanzado permite detectar ataques no conocidos y que por tanto no pueden ser detectados a través de métodos tradicionales.

2.11. Orquestación de la seguridad y automatización de la respuesta ante incidentes mediante SOAR

SOAR es el acrónimo de Security Orchestration, Automation and Response. Es una herramienta con la que nos permite realizar una respuesta ante los incidentes de seguridad más eficaz, más efectiva y manejable a la escala que se necesite implementar.

Lo primero que debemos entender es qué es exactamente la respuesta ante los incidentes de seguridad. A continuación, vamos a enumerar las tres formas actuales de gestionar la respuesta ante incidentes.

- **Respuesta reactiva ante los incidentes de seguridad**. La respuesta ante incidentes de seguridad reactiva permite a los equipos encargados de la seguridad el contener incidentes o ataques, prevenir o controlar daños. A su vez, permite implementar las remediaciones relacionadas con estos. La respuesta reactiva consiste por tanto en mitigar y realizar las remediaciones posteriores a un ataque. Esto se consigue mediante la recuperación de los sistemas afectados, la remediación de todas las brechas de seguridad que generaron el incidente, realización si es necesario de análisis forense de los sistemas afectados, todo lo relacionado con la comunicación y la auditoría posterior para determinar la correcta subsanación del incidente.

- **Threat Hunting**. Este tipo de gestión de los incidentes hace referencia a la búsqueda proactiva de amenazas que pueden ser indicadores de compromisos o incidentes. Estos análisis son llevados a cabo por analista de seguridad expertos dada la complejidad de la acción a acometer. Por lo general, todo esto implica realizar consultas dentro de los datos de seguridad utilizando un SIEM, ejecutar escaneos de vulnerabilidades o pruebas de pentesting contra los sistemas de la organización. Todo ello tiene como objetivo descubrir actividades

sospechosas o anomalías que puedan representar un posible incidente de seguridad.

- **Respuesta proactiva ante los incidentes de seguridad**. Dado que muchos incidentes son descubiertos pasado mucho tiempo, las organizaciones están desarrollando capacidades proactivas de respuesta ante los incidentes que sufren. Esto implica la búsqueda activa en los sistemas monitorizados en busca de algún indicador de compromiso.

Para que un incidente pueda ser subsanado, entra en juego el concepto de gestión o administración del caso o incidente. Esto implica recopilar, distribuir y analizar los datos vinculados a los incidentes de seguridad específicos, para que la respuesta sea efectiva.

Por tanto, esta gestión de casos proporciona las siguientes ventajas:

- Abrir un caso relacionado con un incidente de seguridad confirmado.
- Agregar de forma rápida todos los datos relevantes dentro de una representación digital del caso.
- Se habilita la priorización del caso detectado para poder dar respuesta.
- Permite investigar y agregar información al caso.
- Por último, permite registrar toda la actividad desarrollada posterior al ataque y durante el cierre del caso.

En base a todo lo anterior, los SOAR son una serie de herramientas de seguridad que es definida por Gartner de la siguiente manera:

- Un sistemas que recopila datos sobre amenazas de seguridad y alertas de diferentes fuentes de origen.
- Habilita el análisis de incidentes, el triaje y la priorización, tanto de manera automática como manual siempre con asistencia de la herramienta.
- Permite definir y aplicar un flujo de trabajo estándar para poder desarrollar todas las actividades de respuesta ante los incidentes de seguridad.
- Se codifican los procedimientos de respuesta y análisis de los incidentes de seguridad en un formato de flujo de trabajo digital, esto permite la automatización de algunas o todas las respuestas.

Podemos por tanto considerar que los SOAR poseen tres capacidades clave, que ayudan a los SOC a la labor de gestión de incidentes de manera mucho más efectiva.

- **Orquestación**. Estamos ante la capacidad de coordinar la toma de decisiones y a automatizar las acciones reactivas basadas en la evaluación de riesgos y el estado del caso.

Las herramientas SOAR pueden realizar esto gracias a que se integra con otras soluciones de seguridad de manera que les permite extraer datos e ejecutar acciones proactivas. Los SOAR proporcionan una interfaz genérica, que permite a los analistas definir acciones en las herramientas de seguridad y sistemas IT integradas con el SOAR.

Los SOAR tienen la capacidad de investigar una serie de eventos a través de herramientas internas, también puede manipular, realizar simulaciones y comprobaciones, y enriquecer estos datos, para determinar la relevancia de dichos eventos. Si se confirma que se está ante un incidente, se ejecuta de forma automática un playbook. Dicho playbook que está previamente definido, ejecutará una serie de acciones para solventar el incidente.

- **Automatización**. Relacionada con el anterior punto, se trata de la ejecución de las acciones necesarias para la subsanación del incidente de seguridad dentro de las herramientas de seguridad y sistemas IT integradas dentro del SOAR. Por todo ello, los equipos de seguridad pueden definir los pasos a automatizar.

  Estos pasos como se ha indicado anteriormente se basan en playbooks de seguridad, que los analistas pueden generar a través de por ejemplo una interfaz gráfica.

- **Gestión de incidentes y colaboración**. Mediante esta cualidad los SOAR ayudan a los equipos de seguridad a gestionar estos incidentes de manera más eficiente gracias a la posibilidad de colaborar y compartir datos para poder resolver los incidentes ante los que se enfrentan.

- **Procesamiento de alertas y triajes**. Una herramienta SOAR recopila y analiza los datos de seguridad (generalmente obtenidos del SIEM), correlaciona dichos datos para identificar prioridades y criticidades, y en base a todo ello generar automáticamente incidentes para su investigación. Este incidente provee gran cantidad de información de contexto relevante, esto hace que se pueda investigar el incidente de forma más profunda. Elimina en definitiva la labora humana de identificación de dichos incidentes.

- **Diario y apoyo probatorio**. Los SOAR proporciona un cronograma para poder realizar la investigación. Recopila y almacena artefactos relacionados con el incidente, tanto para el análisis actual como futuro. Se pueden extraer artefactos adicionales para investigar si están relacionados o no con el incidente actual.

- **Administración del caso**. Estas herramientas pueden almacenar y ejecutar acciones y decisiones tomadas por el equipo de seguridad, esto además hace que estas acciones sean visibles para toda la organización, incluso para auditores externos. Con el tiempo, estas herramientas crean una base de conocimiento de la organización a nivel

de amenazas, incidentes y sus respuestas, histórico de las decisiones tomadas ante los incidentes y sus resultados.

- **Gestión de Threat Intelligence**. Estas herramientas aportan datos sobre las amenazas contenidas dentro de las bases de datos abiertas, de líderes de la industria, de organizaciones oficiales o de proveedores comerciales. Los SOAR juntan toda esta información relevante y lo relacionada con incidentes específicos. Todo ello hace que se enriquezca la información sobre el incidente, llegando de forma sencilla hasta el analista de seguridad.

Cuadros de mando e informes.

Los SOAR gracias a su potencialidad se convierten en el eje central de la actividad del SOC. Estas herramientas son capaces de generar informes y paneles que muestran varios niveles de información destinadas a elementos tan dispares como pueden ser los analistas, el mánager del SOC o el CISO. En estos informes se muestran lo eficientes que son a nivel de respuesta ante incidentes, los tiempos que se tardan en gestionarlos o la alineación con otras métricas para determinar el impacto de los incidentes sobre el modelo de negocio.

El SOAR y el SIEM.

Como se ha podido ver a lo largo de todo el capítulo, los SOAR trabajan mano a mano con los SIEM, siendo parte fundamental de todo el entramado del SOC. Los SOAR se aprovechan de su integración con los SIEM para, por ejemplo:

- Recibe alertas y datos de seguridad adicionales para identificar incidentes de seguridad.
- Obtiene los datos necesarios para que los analistas tengan más información a la hora de realizar sus investigaciones.
- Asiste a los analistas en la respuesta proactiva ante un incidente y la búsqueda de amenazas (Threat Hunting).

Los SIEM de próxima generación deben si quieren evolucionar integrar este tipo de herramienta dentro de su core. Si bien los SOAR pueden funcionar de forma independiente, no desarrollan todo su potencial hasta que no se integran junto a un SIEM bien implementado. Por tanto y siguiendo a Gartner, los SIEM deben implementar de forma nativa un sistema para administrar y responder ante los incidentes de seguridad detectados. En fin, una defensa avanzada contra las amenazas dentro de la organización.

2.12. El SOC y la integración con el SIEM.

Un SOC o ISOC es un centro donde el personal de seguridad monitoriza los sistemas empresariales, se gestionan todos los incidentes de seguridad, desde su detección hasta su mitigación. Además, se gestionan todos los riesgos de seguridad posibles.

Dentro del modelo posible de los SOC podemos encontrarnos SOC internos a la organización, servicios externalizados a través de un MSSP, o SOC híbridos en los que se reparten las labores y responsabilidades. Además, surge el concepto de VSOC (SOC virtual) en los que se abandonan el concepto de SOC enmarcado dentro de una instalación física, para ser formado por un grupo de profesionales de la seguridad que trabajan de manera coordinada para realizar las tareas típicas de un SOC.

Por norma general, un SOC posee una serie de funciones diferentes dentro de una organización.

- **Control y análisis forense digital**. Hacer cumplir con todo el cumplimiento de políticas de la empresa, realización de pruebas de penetración (pentesting) para valorar el estado de la seguridad y mejorarla, y la realización de pruebas sobre las vulnerabilidades presentes dentro de la organización.

- **Monitorización y gestión de riesgos**. Se trata de los eventos más relevantes para la seguridad, con el objetivo de identificar posibles incidentes de seguridad y dar respuesta a estos.

- **Administración de las redes y sistemas**. Se trata de la administración de los sistemas y procesos relacionados con la seguridad, nos estamos refiriendo a por ejemplo la administración de identidades, de los accesos a las máquinas, administración de dispositivos de seguridad, …

Aportaciones del SIEM al SOC.

- Los SIEM actuales reducen el ruido blanco (eventos irrelevantes para la seguridad) que se genera dentro de los sistemas corporativos. Junto al SIEM podemos destacar herramientas complementarias basadas en machine learning y analítica avanzada.

- Acceso a todos los elementos de la organización. El SIEM centraliza todos los eventos para hacerlos visibles al equipo del SOC.

- Ayuda a eliminar grandes volúmenes de falsos positivos y la llamada fatiga generada por las alertas. Tradicionalmente, los SOC han generado un gran volumen de falsos positivos en base a la gran cantidad de alertas que se producen. Los falsos positivos generan distracciones a la hora de detectar y gestionar correctamente los incidentes de seguridad. Las nuevas tecnologías permiten precisamente reducir los falsos positivos para que los analistas puedan centrarse plenamente en los eventos y alertas relevantes.

Por todo ello, la tecnología principal de un SOC es el SIEM. Es por tanto su core. Gracias a las funciones descritas a lo largo del capítulo, los integrantes del SOC son capaces de gestionar todo lo relacionado con la seguridad corporativa. Es la herramienta de monitorización plena de la organización.

Alguno de los procesos dentro de un SOC que son facilitados por los SIEM son los siguientes:

- **Investigación de malware**. El SIEM ayuda a combinar datos sobre el malware detectado, correlarlo con el Threat Intelligence y ayudando a mostrar los sistemas y datos afectados. A esto se le puede añadir las nuevas funciones de orquestación para mitigar y solucionar los incidentes de seguridad detectados.

- **Prevención y detección de phishing**. Mediante las correlaciones y el análisis del comportamiento de usuario, se puede detectar si algún usuario ha sido víctima de este tipo de amenaza. Generada una alerta, se puede realizar búsquedas de patrones similares para identificación de nuevas amenazas o compromisos.

- **Investigación de la actividad de insiders**. Los SIEM ayudan a recopilar toda la actividad de un usuario dentro de los sistemas IT de la organización. Dicha búsqueda puede retrotraerse mucho tiempo hacia atrás. Se puede por tanto detectar comportamientos anómalos, exfiltración de datos, …

- **Mitigación de los riesgos generados por despidos de empleados**. Los SIEM pueden identificar intentos de inicios de sesión exitosos o no, de usuarios que ya no están dentro de la organización (ya sea por error, o por acción malintencionada). En concreto, y hasta que se dé de baja una cuenta, se pueden detectar acceso exitosos realizados por exempleados o empleados maliciosos y determinar si ha existido algún tipo de exfiltración de información sensible.

2.13. Comparativa SIEM

La mejor forma de ver y comparar los SIEM actuales del mercado es analizar el Review que realiza la empresa Gartner. Para comenzar vamos a ver el cuadrante mágico de Gartner del año 2018. Destacar cuáles son los SIEM líderes del sector.

Ilustración 2: Cuadrante mágico de Gartner octubre 2018

Dentro del análisis podemos ver que existen a nivel comercial un total de 48 SIEM diferentes. Se nos permite además comparar los SIEM entre ellos en base a sus características.

Vamos a ver las principales características de los principales SIEM del mercado. Estos son Splunk, IBM, LogRhythm, RSA y Exabeam.

Ilustración 3: SIEM líderes

Splunk.

La oferta de Splunk se compone de la versión Enterprise o la versión Cloud. A estos dos modelos, debemos añadirle la capa de seguridad desarrollada dentro de Splunk Enterprise Security (ES). Como componentes adicionales nos podemos encontrar los siguientes:

- **Splunk UBA**. Es el módulo de analítica UEBA que se encarga de todo lo relacionado con este tipo de soluciones.
- **Splunk Phantom**. Módulo con capacidades SOAR.

Splunk Enterprise (o Cloud) se encarga de administrar todo lo relacionado con la gestión de los datos y la capacidad de búsqueda, mientras que la app Splunk Enterprise Security proporciona todas las funcionalidades típicas de un SIEM. Por tanto, se incluyen análisis y paneles en tiempo real, alertas relacionadas con incidentes, gestión de estos, respuesta automatizada (respuesta adaptativa), visualizaciones e informes.

Splunk se puede implementar tanto en modo on-premise como servicio en la nube/IaaS, en modo híbrido y como SaaS a través de Splunk Cloud.

Los elementos de Splunk Enterprise son los siguientes:

- **Indexadores**. Lugar donde se almacenan los datos, transforma los datos desde su formato en raw en eventos y los almacena posteriormente dentro de su índice. También es el lugar donde se buscan los datos una vez indexados.
- **Search Head**. Es la interfaz GUI donde se maneja la administración de las funciones de búsqueda. Se declaran las búsquedas que posteriormente serán enviadas a los peers de búsqueda para ser devueltos los datos y ser mostrados a los usuarios.
- **Universal Forwarder**. Es un agente para la recolección de eventos dentro de una máquina final, y enviarlos a otra instancia dentro de la arquitectura de Splunk, o a un tercero. Este elemento es dedicado, por tanto, únicamente tiene las funcionalidades básicas para la recolección y envío de los logs.
- **Heavy Forwarder**. Muy parecido al anterior, pero con unas mayores capacidades relacionadas con las de los indexadores. Es capaz de parsear datos antes de ser enviados y enrutar tráfico en función del evento. También puede indexar logs antes del envío a otro indexador. Esto lo hace para poder retener logs en caso de caída de la comunicación con el resto de la infraestructura, o bien porque no pueda procesarlo a la velocidad deseada.

Splunk admite una arquitectura flexible y una facilidad a la hora de escalarla. Proporciona a su vez una variedad de aplicaciones complementarias relacionadas con los casos de uso como puede ser Security Essentials, Stream (relacionado con la recopilación de datos a partir de los paquetes de red),

aplicaciones relacionadas con la detección de Ransomware o cumplimiento PCI, … Otra de las grandes ventajas de Splunk, es la existencia de Splunkbase, que es la tienda de aplicaciones que proporciona contenido a toda la arquitectura. Estas aplicaciones pueden ser creadas por la propia Splunk, o por terceros.

La ventaja de Splunk Enterprise es su gran gama de análisis nativa de la que se aprovecha para generar una solución SIEM muy robusta. Posee tanto Splunk ES, Splunk UBA y Splunk Phantom. Splunk ES está por encima del promedio de otros SIEM a la hora del manejo de incidentes de seguridad. Splunk se integra con otros proveedores de seguridad de forma centralizada para gestionar y mejorar la detección de amenazas de seguridad, en lugar de usar otro tipo de herramientas nativas y complementarias. Las integraciones con estos appliances se realiza a través de aplicaciones y complementos que podemos encontrar dentro de Splunkbase. Todo ello haciendo uso de las APIs. Otra app interesante es Splunk Stream relacionado con la recopilación del tráfico de red para determinar la aplicación, le protocolo, incluso si está encriptado. Todo ello es enviado para su posterior análisis en Splunk.

LogRhythm.

Esta solución SIEM la podemos encontrar como LogRhythm Enterprise y LogRhythm XM. Además, podemos encontrar los siguientes complementos:

- **LogRhythm CloudAI**. Módulo que está relacionado con la implementación de UEBA.
- **LogRhythm NetMon**. Se encarga de detectar y responder rápidamente a las amenazas detectadas. Detecta las amenazas en tiempo real, con reconocimiento de aplicaciones, Deep Packet Analytics personalizables, y análisis de comportamiento y de tráfico de red multidimensional.
- **LogRhythm SysMon**. Es un agente instalado en el Endpoint que se encarga de recopilar todos los logs necesarios a nivel de seguridad y cumplimiento y todos los demás que enriquecen la información referente a la actividad de host monitorizado. Puede instalarse tanto en endpoints, como en servidores, y máquinas virtuales tanto de Windows, como de Linux, o Unix.

LogRhythm SIEM se puede implementar como un único dispositivo para soluciones pequeñas (LogRhythm XM) o en formato distribuido para soluciones más grandes (LogRhythm Enterprise).

Para la recopilación de los logs de los sistemas se cuenta con SysMon, para la recopilación de los paquetes de red existe el agente NetMon, y para los datos del core de la plataforma contamos con los colectores (Data Collectors). Por otro lado, los procesadores de datos (Data Processors) se encargan de recopilar y procesar los logs provenientes de los colectores y los agentes SysMon y NetMon, mediante un proceso de análisis, normalización y enriquecimiento. Los indexadores, por último, se encargan de almacenar los logs en formato raw o enriquecidos.

El módulo AI Engine realiza análisis de datos dentro de los Data Indexers. Todo lo relacionado con la administración de los elementos de operación, incluidas las alarmas, la administración de los casos, el flujo de trabajo, las respuestas automatizadas y la administración de la plataforma es realizada por los administradores de la plataforma.

Existe el módulo WebUI para la interacción con la plataforma que puede ser instalado en modo stand-alone o en múltiples instancias. También podemos encontrarnos el CloudAI que es una solución complementaria proporcionada a través de un servicio en la nube que proporciona capacidades UEBA. Toda la arquitectura puede ser implementada en máquinas físicas o virtuales, así como en la nube. Proporciona también soluciones SaaS o por parte del cliente. Permite de forma nativa soluciones de tipo MSPs y MSSP.

LogRhythm ofrece otras características como la monitorización en tiempo real, la priorización de las alertas, … La librería out-of-the-box proporcionada por el módulo AI Engine es extensa y cubre un gran número de fuentes, incluido la telemetría de red (a través de soporte nativo NetFlow o mediante el agente NetMon), así como el CloudAI.

Es una buena solución para aquellos que quieran tener una plataforma única en la que se recojan soluciones SIEM junto a las funcionalidades descritas dentro de NetMon, SysMon y CloudAI. Como desventaja, LogRhythm carece de un store de aplicaciones, y la administración de la plataforma todavía no está integrada dentro de la interfaz de usuario web.

RSA (Dell Technologies).

Esta plataforma denominada NetWitness RSA consta de los siguientes elementos:

- **RSA NetWitness Logs**. Se encarga de proporcionar las capacidades centrales de SIEM.
- **RSA NetWitness Network**. Módulo relacionado con la monitorización en tiempo real de todo el tráfico de red.
- **RSA NetWitness Endpoint**. Relacionado con la monitorización de los dispositivos finales.
- **RSA NetWitness UEBA**. Se trata de la funcionalidad UEBA de RSA.
- **RSA NetWitness Orchestrator**. Solución SOAR con relación OEM con Demisto.

El RSA NWP está orientado a empresas que desean poseer una plataforma completa que proporcione análisis avanzados y en tiempo real para la detección de las distintas amenazas de seguridad. Esto se complementa con el análisis forense de redes y puntos finales, así como la gestión y respuesta de incidentes.

La arquitectura RSA NWP está compuesta por una gran variedad de componentes. Posee gran flexibilidad a la hora de determinar cómo y dónde se instalan los componentes, esto hace que se puedan realizar desde despliegues simples a mucho más complejos, tanto en formato on-premise como en la nube (IaaS o SaaS), así como entornos distribuidos geográficamente.

- **Los decodificadores (Decoders)**. Son utilizados para recopilar logs y datos, así como para realizar analítica con estos.
- **Concentradores (Concentrators)**. Se encargan de la agregación e indexación de los metadatos.
- **Event-Stream Analysis**. Es el módulo responsable del análisis, correlación y notificación de eventos en tiempo real.
- **Archivadores (Archivers)**. Son los encargados de proporcionan la retención de los datos a largo plazo.
- Los **agentes de Endpoint** proporcionan funcionalidades de tipo EDR, así como captura y reenvío de logs.
- **RSA NetWitness Server**. Es la interfaz de usuario (UI) del SIEM.
- **RSA NetWitness Live**. Es un servicio de distribución de contenido basado en la nube.

RSA NWP lo podemos encontrar en una gran variedad de formatos, tanto en implementaciones físicas como virtuales, on-premise o en servicio a través de la nube. Se pueden integrar múltiples fuentes de datos y registros locales, así como fuentes contextuales de datos, y fuentes provenientes de proveedores SaaS comunes y de servicios en la nube.

La solución RSA NetWitness incluye monitorización de los endpoints y de los elementos de red tanto para propósitos de análisis en tiempo real, como análisis forense. Esta tecnología puede considerarse madura dado que se adapta bien a los casos de uso avanzados de defensa contra las amenazas. La contextualización y supervisión del usuario está bien cubierto a través de su solución Fortscale UEBA. Las capacidades de gestión de incidentes son mejoradas a través del módulo SOAR proporcionado por la plataforma RSA NetWitness. El contra de este módulo es que es complejo de implementar y administrar para usuarios menos maduros. Otra desventaja es que todavía no ha adaptado una solución Big Data para la centralización de la recolección de logs. Sin embargo, se mitiga en cierto modo a través del conector nativo denominado Warehouse Connector, que permite el envío de datos a soluciones de tipo Hadoop y NoSQL.

IBM Qradar.

Esta solución SIEM posee los siguientes componentes:

- **IBM QRadar Vulnerability Manager**. Relacionado con la gestión de las vulnerabilidades.

- **IBM QRadar Network Insights**. Módulo que proporciona visibilidad de los flujos de red llamada QFlow.
- **IBM QRadar User Behavior Analytics**. Módulo UBA.
- **IBM QRadar Incident Forensics**. Módulo para el análisis forense.
- **IBM QRadar Advisor** con Watson que proporciona una investigación automatizada de la causa raíz de las amenazas identificadas.
- **IBM** también ofrece una **App Store** donde sus clientes pueden descargarse contenido desarrollado tanto por la propia IBM como por terceros, para ampliar la funcionalidad del sistema.
- **IBM QRadar Network Packet Capture**. Relacionado con capacidades forenses de red más avanzadas.
- **IBM Resilient**. Se trata de una solución SOAR que puede ayudar a las organizaciones a optimizar su proceso de gestión de incidentes de seguridad.

QRadar está disponible como solución en la nube mediante formato SaaS, también está disponible on-premise a nivel de un dispositivo de hardware virtual y como paquetes de software dependientes de los EPS del cliente.

IBM QRadar está pensado para desplegarse en grandes organizaciones, ofreciendo una plataforma robusta para la detección y respuesta antes las amenazas existentes. Aunque preferentemente está preparada para dar soluciones a grandes empresas, también puede implantarse dentro de una organizaciones más pequeñas, adaptándose su contenido al entorno donde se implementa.

QRadar posee una amplia base para poder realizar la implementación y disponibilidad de proveedores de servicios para poder gestionar todo lo referente con la solución de IBM.

Las fortalezas de QRadar según Gartner son las siguientes: Se ofrece un SIEM flexible y potente con un amplio contenido listo para ser usado en una amplia selección de casos de uso. Posee un sólido ecosistema de integraciones que generan un valor añadido con otras soluciones del portfolio de seguridad de IBM, además existe un contenido desarrollado por terceros fácilmente accesible. Gran soporte para la monitorización, con una gran cantidad de firmas para analizar los flujos de datos.

Mientras que las precauciones que han de tomarse son las siguientes: La experiencia de usuario quizá se esté quedando atrás con respecto a nuevas soluciones SIEM del mercado. QRadar exige cierto nivel de madurez en los procesos de seguridad para poder sacar todo el partido a la herramienta. El sistema UBA no está del todo desarrollado. Se detecta que los usuarios dan puntajes más bajos que otros líderes del mercado en relación al servicio que proporciona el fabricante en relación a la integración y despliegue, servicio y soporte.

Exabeam. Exabeam's Security Management Platform (SMP).

Este SIEM está compuesto por seis productos:

- **Exabeam Data Lake**. Almacenaje de los datos basado en Big Data.
- **Exabeam Cloud Connectors**. Módulo para la recolección de eventos provenientes de servicios en la nube.
- **Exabeam Advanced Analytics**. La solución UEBA de Exabeam.
- **Exabeam Entity Analytics**. Relacionado con el análisis de dispositivos conectados a Internet para completar la solución UEBA.
- **Exabeam Threat Hunter**. Módulo relacionado con la búsqueda avanzada de amenazas.
- **Exabeam Incident Responder**. Módulo para la respuesta ante incidentes de seguridad.

Esta solución permite ser desplegada tanto on-premise como en la nube. Dado el gran número de entidades que puede monitorizar, este SIEM está dirigido a grandes organizaciones de sectores como de servicios financieros, atención médica, … Exabeam destaca con respecto a otros competidores por los módulos de Incident Responder o Advanced Analytics. Además, los logs son almacenados dentro de un Data Lake. Exabeam proporciona también el servicio en forma de SaaS.

Como fortalezas detectadas dentro de Exabeam podemos indicar las siguientes: Consta de una arquitectura escalable basada en Elasticsearch y Hadoop (HDFS), además usan Kafka y Spark. A nivel de licencia existe un modelo de fácil entendimiento basado en usuarios y entidades. Funciones de orquestación y respuesta incluyéndose playbooks automáticos para la respuesta ante los incidentes. Proporciona acceso a los datos y al flujo de trabajo de forma granular (predefinidos y personalizables) relacionados con la privacidad. Capacidades de Advanced Analytics, posee una UEBA muy maduro dado que esta fue el origen de su producto antes de entrar en el mercado de los SIEM. Sus clientes destacan de este producto sus altas capacidades en cuanto al servicio de implementación y soporte.

Gartner nos advierte de los siguientes puntos: Las organizaciones con baja madurez dentro de las capacidades de investigación y respuesta ante incidentes obtendrán poco beneficio de este producto. Algunos clientes han indicado que la funcionalidad de respuesta ante incidentes no está tan avanzada en comparación con otros productos del sector. Se ha reportado algún problema referente a respuestas lentas del administrador técnico de la cuenta junto a otros problemas de soporte que han dificultado el crecimiento de la solución SIEM. Exabeam debe todavía implementar el servicio de analítica en la nube dentro del modelo SaaS. Esto limita que sea aplicable en ciertas organizaciones.

Como se puede ver a partir de este pequeño análisis de los SIEM considerados por Gartner como líderes del sector, todos ellos se encuentran en mayor o

menor medida dentro de la considerada tercera generación de SIEMs, donde el almacenamiento de los datos se basa en métodos como Big Data, se usa analítica avanzada mediante por ejemplo machine learning, y para la respuesta ante incidentes se despliegan soluciones de tipo SOAR.

2.14. Búsqueda de amenazas (Threat Hunting) y los sistemas SIEM.

El Threat Hunting es la búsqueda proactiva de amenazas dentro de una organización basándose en la analítica de los eventos contenidos dentro del SIEM. Surge de la necesidad de detectar ataques cada vez más complejos, capaces de evadir las medidas de seguridad tradicionales y que persisten largos períodos de tiempo dentro de la infraestructura afectada. Todo ello se traduce en la existencia de mayores riesgos para las organización. Por tanto, mediante estas acciones lo que se pretende es reducir tanto dicho riesgo, como el impacto posible una vez que se ha producido el compromiso.

La búsqueda de amenazas debe de ser considerado como parte fundamental de los procesos a seguir dentro de una organización para conseguir una protección integral de todos los sistemas.

Este tipo de acciones deben ser efectivas y confiables en sus resultados para lograr los objetivos marcados. Por todo ello, los analistas encargados de realizar estas funciones deben ser capaces de lograr los siguientes objetivos:

- Proporcionar una detección de amenazas precisa y temprana.
- En base a sus detecciones, se debe controlar y mitigar el daño que genera un incidente de seguridad a través de una respuesta temprana.
- Su análisis debe de proporcionar como resultado una mejora de las defensas de la organización para que un posible ataque no tenga éxito.
- Este tipo de actividad hacen que emerjan las debilidades de la organización.

El éxito de este programa tiene dos partes esenciales. La primera debe implicar un proceso formal que debe basarse en una metodología bien definida. La segunda, contribuye a realizar una operación de la búsqueda de amenazas a través de herramientas automatizadas. Los dos puntos anteriores supondrán la mitigación del daño y mejora de la eficiencia global de la seguridad.

Los siguientes son aspectos críticos dentro del Threat Hunting:

- **Personal y sus habilidades**. La integración de las personas, la tecnología y los procesos relacionados con la seguridad son fundamentales. Debe existir una fuerte sinergia entre estos tres elementos.

  En el **Threat Hunting**, la tecnología debe de ser el foco principal, para que las personas pasen a un segundo plano. Aun así, la capacitación del personal es de vital importancia dado que deben de administrar,

configurar e interpretar los resultados proporcionados por dicha tecnología.

- **La esencia entre el Threat Hunting y SIEM**. El objetivo fundamental de esta operativa es la reducción del daño que puede generar un ataque contra la organización. Por ello, las métricas clave deben de ser la mitigación de la exposición ante los posibles ataques, y una mitigación rápida. Estas métricas deben generar un ROI positivo en sus esfuerzos. Este tipo de acciones juegan además un papel crucial en la detección temprana de adversarios y en la eliminación rápida de las vulnerabilidades descubiertas.

- **Valor del tiempo de respuesta**. Se pueden usar tres indicadores principales para confirmar que los esfuerzos desempeñados dan sus frutos:

    - Tiempo de permanencia. Es el tiempo que un adversario ha permanecido dentro de la organización.
    - Reinfección. Determinar cuántos ataques se han lanzado contra la organización.
    - Movimiento lateral. Determina cuánto daño ha causado un adversario, en términos de los sistemas comprometidos.

    Si se tiene una mejora significativa en las áreas anteriores, el programa de Threat Hunting es considerado confiable. Si no es así, hay que replantearlo para su mejora. En todo caso, el SIEM está implementado para realizar la función de defensa reactiva en caso de fallo en el resultado de las metodologías usadas.

- **Métodos de Threat Hunting**. En un programa de búsqueda de amenazas, los analistas utilizan dos métodos básicos, la búsqueda de amenazas basada en el host o en la red. La búsqueda basada en host se utiliza para analizar un sistema individual en búsqueda de los indicadores de ataque (IOA). Por otro lado, la búsqueda de amenazas en la red se utiliza para monitorizar y analizar el tráfico de la red para detectar señales de la existencia de un adversario dentro de la red.

Necesidad de mejora continua. Dado que las amenazas y los adversarios evolucionan, las capacidades de búsqueda y las estrategias de defensa deben a su vez evolucionar también. Para este propósito se debe recurrir a la automatización y la desarrollo de herramientas personalizadas. Estas herramientas deben estar perfectamente alineadas y que sean un complemento de la funcionalidad del SIEM. En el siguiente cuadro podemos ver las diferencias entre el Threat Hunting el Threat Detection según Gartner:

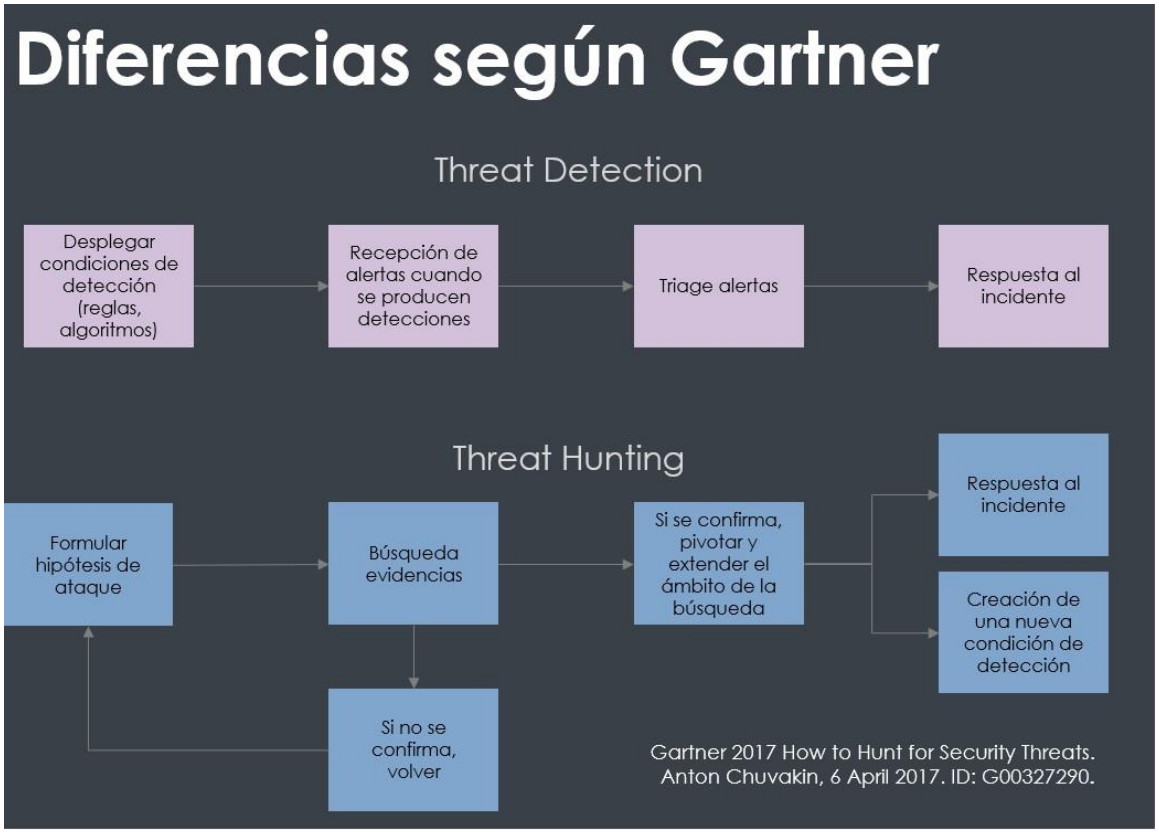

Ilustración 4: Diferencias según Gartner

La eliminación del riesgo al 100% no es posible, siempre puede producirse un incidente de seguridad dentro de la organización, por todo esto, se debe de implementar un doble enfoque de tipo proactivo y reactivo para garantizar lo más posible la seguridad. Por ello, se debe de combinar una solución SIEM junto con el Threat Hunting.

Ciertas plataformas SIEM facilitan este tipo de acciones a través de paneles y acciones creadas a tal efecto. Un ejemplo claro lo podemos encontrar dentro del SIEM de Exabeam. Este SIEM proporciona una interfaz sencilla para simplificar el proceso de creación de consultas complejas, cuando se ha detectado una posible amenaza, el módulo Threat Hunter proporciona líneas de tiempo automáticas en lugar de simplemente los eventos para poder en base a dichas líneas buscar otras otros eventos relacionados de forma más rápida y proactiva. Se ha incorporado a este módulo el marco de trabajo de MITRE ATT&CK. De esta forma, los analistas pueden buscar las tácticas y técnicas entre los distintos usuarios y dispositivos en base a una serie de menús desplegables y una interfaz de seleccionar y buscar. Por último, proporciona basándose por ejemplo en un ID de una alerta o una dirección IP, una línea de tiempo automática en la que pueden aparecer múltiples eventos que nos muestre cómo se ha desarrollado a lo largo del tiempo un posible incidente de seguridad. Con esto se consigue una mayor visión global.

3. Implementación de un sistema SIEM

Para crear nuestro SIEM vamos a usar las dos siguientes soluciones: ELK (proviene de **E**lasticsearch, **L**ogstash y **K**ibana) o como se denomina actualmente, Elastic Stack y Splunk Enterprise (en adelante SE). El Elastic Stack hará funciones de Data Lake, mientras que Splunk Enterprise, aunque también puede hacer estas funciones de almacenamiento masivo, aprovecharemos únicamente sus funciones de SIEM, en particular aprovecharemos su potencialidad a la hora de realizar consultas (lenguaje SPL), la posibilidad de enriquecer los logs, la generación de alertas, reportes y Dashboards.

En primer lugar, vamos a analizar los dos sistemas. Vamos a ver sus características y cómo lo tenemos implementado para que realice las funciones deseadas.

3.1. Elastic Stack.

Para poder entender con qué estamos trabajando, vamos a explicar en los siguientes párrafos qué es exactamente Elastic Stack.

Lo primero que vamos a hacer, es describir la arquitectura de la que se compone.

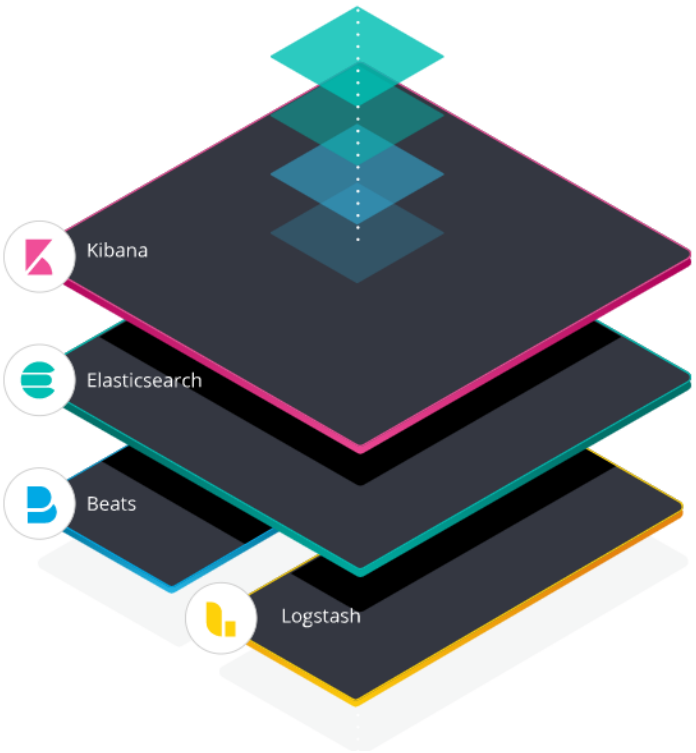

Ilustración 5. Elastic Stack

Como se puede ver en la imagen superior, Elastic Stack está compuesto por varios elementos distintos:

- **Logstash**. Se trata de un pipeline de procesamiento de datos de open source, de lado del servidor que ingesta datos de una gran multitud de fuentes de manera simultánea, transformándolas, para luego enviarlas para su almacenaje.

  Este software ingesta, transforma y envía de forma dinámica los datos. Como se ha indicado, parsea cada evento que recibe, identifica los campos con su nombre para crear la estructura deseada, y posteriormente los trasforma en un formato en común para lograr que puedan ser analizados. Además, Logstash posee una gran variedad de salidas que permiten enrutar los datos donde se deseen.

  Por tanto, su lógica es sencilla. Hay una serie de entradas, unos filtros y unas salidas, que trabajan conjuntamente para formar el pipeline. Se puede resumir como la parte del preprocesamiento antes de ser guardado dentro de Elasticsearch.

- **Beats**. Los Beats son los agentes de datos ligeros. Su único propósito es el envío de datos de múltiples orígenes a los Logstash o Elasticsearch. Existen los siguientes agentes de recolección de eventos:

  - **Filebeat**. Es el agente de recolección de logs y ficheros dentro de la fuente origen.
  - **Metricbeat**. Recopila métricas de los sistemas y servicios donde está instalado. Desde la CPU, a la memoria, … Es por tanto una forma sencilla de enviar estadísticas de sistema y servicio.
  - **Packetbeat**. Monitorea el tráfico de la red. Es un analizador de paquetes de red ligero que envía los datos dentro del core de Elastic.
  - **Winlogbeat**. Recolecta y envía los registros de eventos de Windows.
  - **Auditbeat**. Recopila los datos relacionados con la auditoría de Linux y monitoriza la integridad de sus archivos.
  - **Heartbeat**. Monitoriza los servicios de forma activa. Determina por tanto si están arriba o no. En base a una lista de direcciones, este beat lo que hace es determinar si está activo o no.
  - **Functionbeat**. Se implementa dentro del proveedor de la nube donde se aloja el servicio a monitorizar (FaaS – Function-as-a-Service), recopilando, enviando y monitorizando los datos de los servicios alojados en la nube.

  Una de las grandes ventajas con las que cuenta este sistema, es que se pueden crear Beats de manera personalizada y puesta a disposición de la comunidad. Esto hace que se enriquezcan los eventos recolectados y la forma de hacer esto.

- **Elasticsearch**. Es el motor de búsqueda y análisis distribuido dentro del core del Elastic Stack. Este motor de búsqueda está basado en Apache

Lucene. Permite indexar y recuperar documentos de tipo JSON (sin esquema rígido) en distintos formatos. Es decir, está orientado a documentos. Está basado en Java y proporciona una interfaz API RESTFUL de acceso. Esta licenciado bajo las condiciones de la licencia Apache.

Posibilita búsquedas en tiempo real para todo tipo de datos. Estas búsquedas pueden ser estructuradas como no estructuradas. Elasticsearch es de naturaleza distribuida (escalado de forma dinámica, implementando la HA), lo que hace que pueda crecer sin problemas a medida que los datos almacenados lo van haciendo. Permite por su arquitectura una gran velocidad a la hora de obtener los datos a través de las consultas.

- **Kibana**. Es una herramienta que permite visualizar y explorar los datos que se encuentran indexados dentro de Elasticsearch. Por tanto, se trata del GUI de usuario que permite obtener de forma gráfica resultados basados en la analítica de los datos contenidos dentro de Elasticsearch. Kibana posee las siguientes características:

    - **Exploración y visualización**: Visualizaciones como pueden ser tablas, métricas, Canvas, … Dentro de la exploración de datos podemos ver funciones como la de Discover, Dashboards, … Existen también Dashboards preconfigurados, elementos relacionados con el Machine Learning y con otros elementos de tipo Compartir y Colaborar.
    - **Administración y monitorización**: Dentro de estas funcionalidades podemos encontrarnos con una gran cantidad de funcionalidades. Elementos de seguridad (SSO, RBAC, …). Administración de elementos como son los índices y su ciclo de vida, de la licencia, … Relacionado con la monitorización existen elementos como la monitorización completa de salud del Stack o la configuración de la política de retención. Se gestiona todo lo relacionado con las alertas. También podemos encontrarnos con las herramientas para desarrolladores y todo lo relacionado con el despliegue tanto en la Cloud como por ejemplo con los Dockers.
    - **Soluciones**. Dentro de esta categoría podemos encontrarnos con Elastic Maps, Logs, Infraestructure, Uptime, APM y, sobre todo, Elastic SIEM.

    En todo caso, lo descrito en los puntos anteriores es tan solo un pequeño resumen de las grandes capacidades que posee Kibana. Si se desea, se pueden ver todas las funcionalidades en el siguiente enlace.

Arquitectura implementada.

Aquí se describe la arquitectura implementada dentro del SOC a nivel de SIEM (uso de Splunk) y Data Lake (uso de Elastic Stack).

Para la detección y generación de los eventos relevantes (EN) para la seguridad y que posteriormente se enviarán al SE, vamos a usar dos motores de búsqueda distintos. ElastAlert y Watcher.

ElastAlert.

Este correlador es un framework desarrollado para generar alertas en base a anomalías, spikes u otros patrones que puedan ser de interés dentro de los datos contenidos en Elasticsearch. Este correlador está desarrollado por Yelp. ElastAlert está escrito en Python. Posee licencia Apache License, Version 2.0.

Su lógica es la de cualquier sistema de alertas, realiza una serie de consultas al Elasticsearch y en función del resultado de ésta, lanzar una acción determinada.

Está diseñado buscando la confiabilidad, que sea altamente modular y fácil de configurar. Dentro del ElastAlert podemos encontrar un conjunto de reglas a usar, cada una de estas reglas definen el tipo de consulta que se va a realizar. Existen varios tipos de reglas:

- **Frequency**. De tipo frecuencia. Se dispara la alerta cuando hay al menos X eventos coincidentes dentro de un tiempo Y.

- **Spike**. Alerta de tipo pico. La alerta se dispara cuando la tasa de eventos aumenta o disminuye un valor determinado que se le ha indicado previamente.

- **Flatline**. Alerta que se genera cuando hay menos de X eventos dentro de una ventana de tiempo Y.

- **Blacklist/whitelist**. Estas alertas se generan cuando un determinado campo coincide con un valor determinado dentro de este tipo de listas.

- **Any**. Son alertas que se disparan cuando coincide cualquier evento con un filtro determinado.

- **Change**. Se disparan cuando un campo determinado posee dos valores diferentes dentro de un tiempo dado.

- Tipo **new_term**. Se dispara cuando un término nunca antes visto, aparece en un campo determinado.

- **Cardinality**. De tipo cardinalidad. Se dispara cuando el número de valores únicos para un campo determinado está por encima o por debajo de un umbral indicado previamente.

Posee además un gran número de conexiones con otras aplicaciones para la notificación de las alertas: email, JIRA, Telegram, VictorOps, …

Cada uno de estos tipos de alertas podrán ser usadas para la generación de los eventos notables.

Existe una amplia documentación oficial de las funcionalidades de la herramienta.

ElastAlert está implementado dentro de nuestra solución en un servidor independiente a la infraestructura del Elastic Stack, en alta disponibilidad.

Watcher.

Watcher es una funcionalidad que se encuentra implementada dentro del Elastic Stack y que es totalmente gratuita. Se usa para la creación de acciones basadas en una serie de condiciones (búsquedas filtradas), que se evalúan de forma periódica mediante la consulta de los datos almacenados dentro del núcleo. Esta definición no es más que la generación de búsquedas planificadas en el tiempo y que cuando cumplen la condición dada, generan una acción. En este caso la acción será enviar la información de la forma deseada a la arquitectura de Splunk.

Existe como todo lo relacionado con Elastic una amplia documentación que se debe consultar para saber la forma en que funciona y cómo debemos realizar las cosas para conseguir que se envíen los eventos detectados.

Actualmente podemos crear alertas de dos formas distintas, podemos crear un watch de manera simple a través de un entorno gráfico, pero si se desea crear un watch más complejo, se puede hacer a través del lenguaje DSL (Domain Specific Language) basado en JSON. Este lenguaje es un AST (Abstract Syntax Tree) de consultas, que posee dos tipos de cláusulas:

- **Cláusula de consulta de hoja**. Estas cláusulas buscan un valor particular dentro de un campo determinado. Estas coincidencias pueden ser por los siguientes tipos:

    - **Match query.** Devuelve documentos cuando coinciden con un texto, número, fecha o valor booleano proporcionado.
    - **Term query**. Nos devuelve un documento que contiene un término exacto en un campo proporcionado. Se usan para la obtención de documentos que poseen un valor preciso.
    - **Range query**. Devuelve documentos que contengan términos dentro de un rango previamente proporcionado.

- **Cláusula de consulta compuesta**. Estas cláusulas encapsulan otro tipo de cláusulas hoja o compuestas, y son usadas para combinar múltiples consultas de manera lógica (consultas tipo must, filter, should o must_not dentro del tipo bool, o la de tipo dis_max) o para alterar su comportamiento (como son las consultas constant_score). Estas consultas se comportan de manera diferente si son usadas dentro de un contexto de consulta o en uno de filtro.

Esto último se explica de la siguiente manera. Elasticsearch ordena los resultados de las búsquedas en base a una serie de puntajes de relevancia, esta puntuación nos dice cómo de coincidente es cada documento devuelto en base a la consulta realizada. Este es un número de tipo coma flotante, que cuanto mayor es su valor, más relevante es el documento. Este valor la ser calculado depende de si la cláusula de consulta se encuentra dentro de un contexto de consulta o de filtro:

- **Contexto de consulta**. En este contexto se decide si el documento coincide o no con lo deseado. Se calcula además una puntuación de relevancia. El contexto de consulta está vigente cada vez que se pasa una cláusula de consulta a un parámetro query. Son por tanto las búsquedas en sí (parámetros dentro del search API).
- **Contexto de filtro**. Se utiliza principalmente para filtrar datos estructurados. Este tipo de contexto estará vigente siempre que se pase una cláusula de consulta a un parámetro filter (ejemplos como los parámetros filter o must_not dentro de los parámetros de tipo bool).

Podemos encontrar en la siguiente documentación oficial todo lo que necesitemos para entender el proceso de creación, gestión, activación, … de los watches.

Durante el desarrollo del trabajo se irá explicando en detalle los elementos usados para la generación de los distintos eventos notables. A continuación, una vez que hemos visto la parte relacionada con el Data Lake, pasamos a describir todo lo relacionado con la solución elegida como SIEM.

3.2. Splunk.

El objetivo último de implementar la lógica del SIEM dentro de Splunk es conseguir una serie de funcionalidades que si lo intentamos realizar dentro de Elastic Stack no podríamos. Por tanto, dentro del SE podremos conseguir realizar lo siguiente:

- Centralizar todos los eventos de seguridad dentro de un único índice. El número de eventos que contendrá dicho índice será muy inferior al que se contendría en caso de no existir este tipo de filtrado. Como se ha comentado anteriormente, el Data Lake será la parte del Elastic Stack, y el SIEM como tal, será el SE.

- Eventos notables o relevantes para la seguridad basados en etiquetas en base a una serie de criterios preestablecidos. Estos valores, junto al que define su identidad, serán los que se usen como criterio de búsqueda y correlación.

- Reducción del ruido y de los falsos positivos en base a un sistema de scoring según se trate de una identidad dada, o en la fase del Ciber Kill Chain en la que se encuentre.

- Creación de alertas o casos de uso basados en una gestión de identidades y de pesos a la hora de dispararse una alerta.

- Generación de Dashboards y Reports para su análisis. Mediante este tipo de elementos se podrán realizar análisis y detección de anomalías que no sean cubiertos por los distintos casos de uso implementados.

- Búsqueda de patrones y anomalías en base a consultas históricas de los registros almacenados. Dado que se dispondrá de un histórico de eventos con una retención muy amplia, se podrán realizar búsquedas para detectar patrones o anomalías generadas a lo largo de amplios períodos de tiempo que por su naturaleza las alertas de seguridad no pueden cubrir.

Arquitectura de Splunk.

Para entender los elementos de los que consta la arquitectura de Splunk, debemos entender el ciclo de vida de los logs. Existe una entrada, un almacenamiento y una parte relacionada con la búsqueda.

Por tanto, cada elemento de la arquitectura de SE estará orientado a una de estas etapas.

Entrada de datos.

Es la etapa en la que se recolectan los logs provenientes Están constituidos por los Forwarders:

- **Universal Forwarder**. Recopila los logs de las fuentes de datos. Estos datos los puede enviar a otro Forwarder o directamente a los indexadores. Utiliza menos recursos hardware que los Heavy, siendo también mucho más escalable dado que se pueden instalar em miles de clientes finales. Se pueden instalar en múltiples plataformas como pueden ser Windows o Unix. Estos agentes incluyen únicamente los componentes esenciales para el reenvío de datos a otras instancias. Pueden ser configurados in-situ, o a través del Forwarder Management.

- **Heavy Forwarder**. Es el otro recopilador y el encargado de enviar los logs a la arquitectura de Splunk (incluso puede enviar los logs a un tercero). Posee casi todas las características de un indexador, pero, por ejemplo, no puede realizar búsquedas distribuidas. Su gran característica diferenciadora frente al Universal es que analiza los datos antes de reenviarlos, y puede enrutarlos a diferentes sitios, según criterios de origen o de tipo de evento. También tiene la funcionalidad de

indexar los datos de forma local, mientras reenvía los datos a otro indexador. Por tanto, se usará este tipo de solución cuando sea necesario procesar datos no estructurados antes de su envío.

Se encargan de recolectar los datos y enviarlos dentro del core de la infraestructura de Splunk (o a un tercero), para su indexación y búsqueda.

Por último indicar, que se podrán indexar eventos de otras formas. Una de ellas será precisamente la forma en que se van a enviar los eventos notables desde el Data Lake a Splunk. No estamos refiriendo al método denominado como Webhook o en la nomenclatura de Splunk, HTTP Event Collector. Desde los dos correladores enviaremos los datos de los eventos notables a través de un HTTP POST hasta el Search Head.

Almacenamiento de datos.

La etapa del almacenamiento de los datos consta de dos fases: análisis e indexación.

En la fase de análisis, el software lo que hace es examinar, analizar y transformar los datos para extraer únicamente la información relevante. Todo esto se conoce como el procesamiento de los eventos. Es en este momento cuando el flujo de datos es particionado en eventos individuales. Esta fase consta de la siguientes fases:

- Rompe el flujo de datos en datos individuales.
- Identifica, analiza y establece las marcas de tiempo.
- Se generan los valores de tipo metadatos de los distintos eventos.
- Se pueden aplicar transformaciones de datos dentro de los eventos en base a expresiones regulares.

Indicar, que esta fase también puede llevarse a cabo en la etapa anterior dentro de los Heavy Forwarders.

En la etapa de indexación, Splunk escribe los eventos dentro de los indexadores en base a sus índices. Escribe los datos sin procesar de forma comprimida. El elemento dentro de la arquitectura que se encarga de realizar este proceso es el indexador.

- **Indexer**. Es una instancia de Splunk que se encarga de indexar los datos, transforma eventos sin procesar (si no se ha realizado con anterioridad) y los guarda dentro de su índice correspondiente. También se encarga de buscar los datos indexados en respuesta a las distintas solicitudes de búsqueda realizadas. En arquitecturas distribuidas, es el Search Head el que realiza la búsqueda distribuida. Por tanto, dentro de los indexadores tenemos datos sin procesar en forma comprimida, índices que apuntan a dichos datos (tsidx), más algunos archivos de metadatos. Los datos se almacenan dentro de contenedores denominados buckets.

Una de las grandes ventajas que supone la creación de un clúster de indexadores es que los datos quedan replicados. Esto hace que, si un nodo del clúster cae, el dato siga siendo consultable.

En todo este proceso se han introducido dos conceptos fundamentales dentro de lo que es una arquitectura distribuida: búsqueda distribuida y clúster de indexadores.

- La **búsqueda distribuida** surge cuando en una arquitectura se divide la gestión de búsqueda, de las actividades de indexación. En este tipo de escenario, son los Search Head los encargados de realizar las búsquedas a través de otras instancias (indexadores). Estas instancias son denominadas search peers, y que son los que realmente realizan la búsqueda. Es el Search Head el que une los resultados para mostrárselos al usuario.

  Este tipo de búsqueda provee una escalabilidad horizontal, una instancia por tanto puede dividir su función indexando y buscando eventos a la vez gran cantidad de datos. Además, permite correlar eventos pertenecientes a silos de datos.

- Un **clúster de indexadores** lo que permite como se ha indicado anteriormente es la posibilidad de tener los datos replicados. Esto hace que se implemente una sistema en alta disponibilidad y una rápida recuperación ante desastres. Cada uno de los elementos del clúster son denominados como peer node.

Búsqueda de datos.

Dentro de una arquitectura distribuida, podemos encontrarnos con la situación que haya una gran cantidad de datos a consultar e indexándose, y múltiples usuarios realizando consultas. Esto justifica que el indexado recaiga en máquinas dedicadas, y que las búsquedas lo hagan sobre otro tipo de instancia. Es aquí donde entra en juego el Search Head. El SH proporciona una interfaz gráfica de usuario con la que éste interactúa con el sistema. De este modo, el usuario puede realizar búsquedas sobre los datos. Si únicamente realiza funciones de búsqueda, se denomina SH dedicados.

El SH se encarga de enviar las consultas a los indexadores, son en estos últimos donde realmente se realizan las búsquedas. Es en el SH donde se combinan los datos devueltos, para ser mostrados al usuario (son las búsquedas distribuidas que se ha comentado anteriormente).

También existe el concepto de clúster de Search Head. Son una serie de SH (mínimo tres) que se coordinan para realizar las búsquedas. Asignan según la carga los distintos trabajos, garantizando que todos ellos siempre tengan acceso a los mismos objetos de conocimiento (knowledge objects).

Otros elementos.

Dentro de los elementos de la arquitectura distribuida de Splunk podemos encontrar otros elementos clave como son los siguientes: License server, Deployment server, Cluster master.

- **License server/master**. Es el encargado de controlar una o más licencias esclavo. Es el encargado de controlar todo lo relacionado con la gestión de la licencia de Splunk dentro de toda la arquitectura.

- **Deployment server**. Se trata de una instancia de Splunk que hace las veces de administrador a nivel de configuración centralizada, agrupando y administrando de forma colectiva cualquier otra instancia dentro de SE. Los elementos que son administrados por esta instancia son denominados deployment clients. El DS se encarga de mantener a los clientes actualizados a nivel de configuración, aplicaciones, … estas unidades que se despliegan se denominan deployment apps. Una forma de gestionar todo de forma gráfica y fácil es a través del forwarder management interface.

- **Cluster master/master node**. Se trata del encargado de regular el comportamiento del clúster de indexadores. Se encarga de coordinar la replicación de los peer nodes, y le dice al SH dónde encontrar los datos. También ayuda a administrar la configuración de dichos peers nodes y a realizar acciones correctoras en el caso de que uno de los nodos pierda conectividad. Estas instancias no indexan datos, existiendo un único maestro por clúster.

Por tanto, la arquitectura que tenemos implementada es una arquitectura on-premise (Splunk Enterprise), que consta de un SH, un HF, un clúster formado por dos indexadores que generan la alta disponibilidad de los datos, y una máquina donde se alojan el ML, MN y el DS. Indicar, que también se indexarán eventos de forma directa contra Splunk (por tanto, sin pasar por el Data Lake) en base a aprovechar la licencia contratada, y las posibilidades a nivel de integración directa que proporciona Splunk junto a otras soluciones de seguridad.

Dentro de la documentación de Splunk existe una amplia gama de documentos que profundiza en estos conceptos. El siguiente enlace nos dirige al índice de la documentación de Splunk. Des de aquí, podemos buscar información sobre cualquier temática.

3.3. Lenguaje SPL.

Uno de los principales motivos de elegir a Splunk como solución SIEM fue la gran potencialidad y a la vez sencillez, que proporciona su lenguaje de consulta denominado SPL. Veamos a continuación algo de teoría sobre dicho lenguaje.

SPL (Splunk's Processing Language). Es el lenguaje de consulta de Splunk. Combina las mejores capacidades del lenguaje SQL con el uso de los pipeline de Unix permitiendo acceder a todos los datos en su formato original, optimizar los eventos de series temporales y usar el mismo lenguaje en las visualizaciones.

SPL proporciona más de 140 comandos que permiten buscar, correlacionar, analizar y visualizar cualquier dato. Este lenguaje se puede resumir en cinco áreas clave:

- **Búsquedas muy especialidades**. Busca por palabras clave y filtra a través de cualquier conjunto de datos. Se pueden realizar búsquedas más complejas utilizando sub-búsquedas.

- **Enriquecer y explorar**. Se puede usar el comando "lookup" para unir datos no estructurados con otros que sí lo son. Se pueden usar los comandos "cluster" y "analyzefields" para encontrar predictores de campos y las relaciones entre distintos conjuntos de datos.

- **Visualización de datos geográficos en tiempo real**. Utilizando el comando "iplocation" asignado a las direcciones IP, se puede obtener la latitud y longitud, y usando el comando "geostats" se pueden mapear estadísticas en tiempo real.

- **Predecir, realizar gráficos y visualizaciones estadísticas**. Mediante el comando "stats" que posee más de 20 opciones diferentes, se pueden calcular estadísticas y generar tendencias. Mediante estos valores, podemos generar visualizaciones gráficas en cualquier rango de tiempo y con granularidad.

- **Aprendizaje automático y detección de anomalías**. A partir de las anomalías detectadas, podemos descubrir actividades y eventos inusuales. Mediante los comandos "fit" y "apply" se pueden crear y aplicar modelos de aprendizaje automático.

Para más información sobre las funcionalidades del lenguaje SPL podemos recurrir a los siguientes recursos:

- Exploring Splunk: Search Processing Language (SPL) Primer and Cookbook.
- Tutorial de búsqueda.
- Manual de referencia de búsqueda.
- Curso propio de Splunk sobre la materia.
- Además, existen múltiples vídeos que ayudan en su comprensión:
    - Basic Searching in Splunk.
    - Splunk & Machine Learning.

- Y un largo etcétera de vídeos y canales donde se explican en detalle cada una de las funcionalidades que poseen los comandos de Splunk.

A lo largo del desarrollo de las alertas se irán describiendo las distintas funcionalidades de las consultas que vayamos implementando.

3.4. Normalización CIM (Common Information Model).

Uno de los elementos principales para una correcta correlación es que los campos de los distintos eventos tengan el mismo nombre cuando hacen referencia al mismo significado. Una IP de origen puede tomar distintos valores dependiendo de la fuente de origen (src_ip, src, srcIp, …) Esto hace que cuando deseamos buscar eventos distintos relacionados con una misma identidad, o no se pueda, o se deban contemplar todo el abanico de posibilidades dentro de la consulta. Para evitar todos estos problemas, se introduce el concepto de normalización de los campos dentro de los eventos. A partir de la normalización el campo IP origen pasa a llamarse, en cualquier caso, por ejemplo, src_ip.

La normalización que hemos elegido y que se puede aplicar tanto en origen (que será lo que haremos) como dentro de la arquitectura de Splunk, es la que se implementa dentro de Splunk.

Otro motivo justificado para usar esta normalización es que en un futuro con los campos en formato CIM se podrá migrar más fácilmente a una solución como es Splunk Enterprise Security.

Veamos pues en qué consiste el modelado CIM. Es un modelo semántico compartido cuyo objetivo es extraer valor de los datos. Se busca por tanto en base a una colección de modelos de datos, el tratamiento coherente y normalizado de los datos para conseguir una máxima eficiencia en el momento de la búsqueda.

Cada modelo de datos consta de un conjunto de nombres de campo que definen el mínimo común denominador de un dominio de interés (por ejemplo, la IP de origen que se comentaba anteriormente).

Mediante datos modelados correctamente podemos relacionar distintas fuentes, acelerar los datos clave en búsquedas y paneles, o la creación de informes y visualizaciones. Otro elemento que también usaremos serán las etiquetas. Estas etiquetas preconfiguradas permiten también agrupar valores de campo relacionados por el modelo de datos (crearemos etiquetas personalizadas, junto a las propias que ya existen dentro del modelo CIM). Todo lo anterior permitirán que los análisis, validación y las alertas sean implementadas de forma más fácil y consistente.

El motivo de la existencia del modelado CIM es que ayuda a la normalización de los datos dentro de Splunk. Estos datos coinciden con un estándar común,

utilizan los mismos nombre de campo y etiquetas agrupadas por eventos relacionados de diferentes fuentes o proveedores. Además, otra de las grandes ventajas que introduce esta normalización es la posibilidad de la aceleración de datos. Es interesante el parar de analizar el modelado CIM por un momento para explicar en qué consiste la aceleración de los datos: lo que se acelera es un modelo de datos, con esta acción lo que se le está diciendo a Splunk es que resuma sus datos sin procesar periódicamente, haciéndolo únicamente sobre los campos indicados. Es decir, se reindexan grupos de datos con un menor tamaño (solo los campos necesarios normalizados). Esto tiene un primer coste negativo a nivel de rendimiento y de consumo de licencia, pero que posteriormente se ve compensado en cuanto se pueden realizar búsquedas mucho más rápidas y por tanto más eficientes.

El CIM actúa como un esquema de tiempo de búsqueda ("esquema sobre la marcha") para permitirle definir relaciones en los datos del evento mientras deja intactos los datos sin procesar de la máquina.

Lo normal dentro de una arquitectura dedicada de Splunk es que los datos de las fuentes provengan de forma ‘desordenada’ de las distintas fuentes de origen. Dichos datos se almacenan en bruto (raw), para posteriormente y ya en tiempo de consulta, y mediante el uso de los add-ons, normalizase los campos. Sin embargo, en el presente proyecto la mayoría de las veces no lo haremos así. En base a la normalización que conocemos, enviaremos los eventos de origen (es decir, desde el Data Lake) ya normalizados. De esta forma, no será necesario ningún tipo de add-on (que deberíamos crear) de parseado para la normalización. De hecho, dado que los campos que provienen del Elastic son campos con valores tipo JSON, no tendría demasiado sentido el tener que crear los ficheros de parseo (transforms y props) dentro del SH en base a expresiones regulares cuando ya desde el origen le podemos dar el valor de campo deseado.

Actualmente dentro del modelo CIM de Splunk podemos encontrar hasta un total de 22 categorías. Estas son las siguientes: Alerts, Authentication, Certificates, Change, Databases, Data Loss Prevention, Email, Endpoint, Interprocess Messaging, Intrusion Detection, Inventory, Java Virtual Machines (JVM), Malware, Network Resolution (DNS), Network Sessions, Network Traffic, Performance, Splunk Audit Logs, Ticket Management, Updates, Vulnerabilities y Web.

Cada uno de los eventos que enviemos desde el Data Lake a Splunk deberá de identificarse dentro de cada una de estas categorías, para posteriormente identificar el nombre de cada uno de los campos de interés.

Como ejemplo vamos a ver cómo se parsearía un evento de tipo Malware.

Dentro de la documentación de Splunk podemos ver que las etiquetas (tags) asociadas a este tipo de evento son las siguientes:

Dataset name	**Tag name**
Malware_Attacks	malware

	attack
Malware_Operations	malware
	operations

Tabla 4. Etiquetas para tipo Malware

La documentación de Splunk nos indica que Malware_Attacks está diseñado principalmente para buscar y crear alertas relacionadas con posibles infecciones de malware en el entorno analizado. El Dataset Malware_Operations está relacionado con la monitorización de la salud y del estado operativo de las soluciones antivirus.

Nos facilita a su vez, una tabla donde se enumeran los campos extraídos y calculados para este Dataset en concreto. Un simple ejemplo lo podemos encontrar en parte de la siguiente tabla:

Nombre del Dataset	Nombre del campo	Tipo de dato	Descripción	Lista de posibles ejemplos
Malware_Attacks	action	string	Es la acción que realiza el dispositivo.	ES esperado: allowed Otros: blocked, deferred
Malware_Attacks	category	string	Es la categoría del evento de malware, por ejemplo, podría ser un keylogger o un add-ware.	

Tabla 5. Ejemplo Dataset Malware

Por tanto, lo que vamos a tener que realizar es determinar los campos de los eventos relacionados con este tipo de categoría, y determinar si existe su equivalente. De ser así, dicho campo se normalizará. En caso de encontrarse con un campo no normalizado, se analizará si es posible realizar una normalización ad-hoc.

3.5. Eventos notables.

Este es un concepto adoptado de Splunk. Podemos llamarlo de este modo, o llamarlo por ejemplo evento destacado, evento a seguir relacionado con la seguridad, … En todo caso, el concepto no cambia. Los eventos notables son eventos que bajo unas condiciones determinadas suponen ser lo suficientemente relevantes como para que sea necesario una investigación. En nuestro caso, vamos a utilizar este concepto para crear los eventos relevantes dentro del ámbito de la seguridad. Los eventos notables que se van a crear en el presente trabajo serán eventos de seguridad relevantes, pero con el matiz de

que no todos generan necesariamente una alerta dentro del SIEM, esto se producirá en base a un peso determinado.

Las correlaciones que generarán los eventos notables no son más que las clásicas alertas que generan una acción. Por tanto, un evento notable es una alerta (y una alerta no es más que una búsqueda planificada en el tiempo que como salida posee una acción determinada) que tiene como salida la generación de un evento (un nuevo evento con una serie de campos normalizados) dentro de un índice especial. Además, estos eventos se podrán analizar dentro de su propio dashboard en Splunk.

Los eventos notables poseen de por sí una severidad determinada. Esta severidad estará en función tanto de la propia naturaleza en sí del evento generado, como de la prioridad del asset. Por tanto, se debe tener listado todos los posibles assets que se van a monitorizar junto a la importancia de cada uno de estos dentro de la organización. Como se ha dicho, la otra variable para determinar la severidad dependerá de lo que signifique contextualmente la alerta. No es lo mismo la detección de un evento de login fallido sobre una cuenta, que la misma acción precedida por un login satisfactorio. En la primera el ataque no ha tenido éxito, mientras que en la segunda sí. Esto hará lógicamente que su prioridad a la hora de su tratamiento cambie totalmente.

Para poder determinar la severidad de cada una de las alertas vamos a implementar un sistema de calificación de cada uno de los eventos notables que se van a implementar en base a los siguientes criterios:

- Posición del evento dentro del Cyber Kill Chain (CKC).
- El éxito o no de la acción que se está detectando.
- El número de assets implicados en la detección.

En base a estos valores, junto al propio contexto del asset, tendremos la prioridad inicial de la alerta. Este valor determinará, por tanto, la urgencia con la que se debe gestionar un evento determinado.

Como ya conocemos, el Cyber Kill Chain consta de las siguientes etapas: Reconnaissance, Weaponization, Delivery, Exploitation, Installation, Command & Control, Action on Objective.

Una forma de poder relacionar las alertas con la fase del CKC es aplicar la siguiente tabla:

Phase	Detect	Deny	Disrupt	Degrade	Deceive	Contain
Reconnaissance	Threat Intelligence NIDS D/B Security	Information Sharing Policy				
Weaponization	Threat Intelligence NIDS					
Delivery	Context-Aware Endpoint Malware Protection	Change Management File Integrity Application Whitelisting NIPS	Inline AV	Queuing		Router ACLs App-Aware Firewall Trust Zones Inter-Zone NIPS
Exploitation	Endpoint Malware Protection	Secure Password	DEP			App-Aware Firewall Trust Zones Inter-Zone NIPS
Persistence / Lateral Movement	Log Monitoring	Privilege Separation Secure Password Two- Factor	Router ACLs AV			App-Aware Firewall Trust Zones Inter-Zone NIPS
Command & Control	NIDS	Firewall ACL	NIPS	Tarpit	DNS Redirect	Trust Zones DNS Sinkholes
Actions on Targets	Endpoint Malware Protection	Encryption	Endpoint Malware Protection	Quality of Service	Honeypot	Incident Response
Exfiltration	DLP	Egress Filtering	DLP			Firewall ACLs

Ilustración 6. Fases CKC

En esta tabla se puede ver cómo en base del dispositivo de seguridad de donde provenga el evento generado, podremos determinar en qué fase estamos dentro del CKC.

En referencia al CKC indicar que han existido investigaciones críticas que han formulado modificaciones y adaptaciones dentro de ésta. De todas ellas destaca la llamada Unified Kill Chain. En esta versión, se unifican el clásico Cyber Kill Chain con el framework de MITRE ATT&CK. Ésta contiene un total de 18 fases de ataque únicas que cubren todos los posibles supuestos dentro de un ciberataque. Está pensada por tanto para analizar, comparar y defenderse de las posibles APTs. A continuación, podemos ver las distintas fases del CKC unificado.

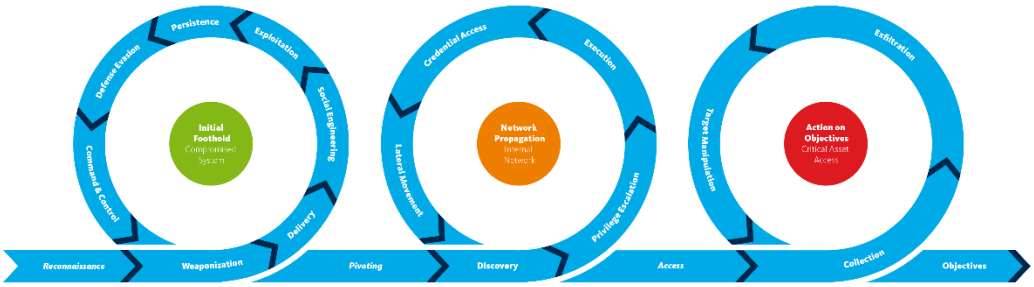

Ilustración 7. Unified Kill Chain

Los 18 elementos de este Cyber Kill Chain unificados son los siguientes:

<u>Punto inicial de compromiso del sistema (Initial foothold).</u>

- Reconocimiento (Reconnaissance).
- Armado (Weaponization).

- Entrega (Delivery).
- Ingeniería social (Social Engineering).
- Explotación (Exploitation).
- Persistencia (Persistence).
- Evasión defensiva (Defense Evasion).
- Comando y control (Command & Control).

Al siguiente nivel se llega a través del Pivoting.

Propagación por la red (Network propagation).

- Descubrimiento (Discovery).
- Escalada de privilegios (Privilege Escalation).
- Ejecución (Execution).
- Acceso a través de credenciales (Credential Access).
- Movimiento lateral (Lateral Movement).

Una vez que se produce el acceso dentro del sistema tenemos la siguiente fase.

Acción sobre los objetivos (Action on objectives).

- Colección (Collection).
- Exfiltración (Exfiltration).
- Manipulación de los objetivos (Target Manipulation).
- Objetivos (Objectives).

Todas estas diferenciaciones a nivel de categoría son muy interesantes, pero para el desarrollo inicial del trabajo se considera que con aplicar el modelo tradicional del CKC es suficiente para realizar una categorización de prioridad adecuada.

Por último, vamos a ver en qué consiste el framework de MITRE ATT&CK (Adversarial Tactics, Techniques and Common Knowledge) del que algo ya hemos comentado algo con anterioridad.

MITRE ATT&CK es una manera de describir y categorizar los comportamientos de los adversarios. Por adversarios se entiende cualquier actor que pueda realizar un ataque contra una organización. Podríamos considerarlo como una base de datos de tácticas, técnicas y procedimientos usados en la vida real. Basado en una serie de matrices, permite describir la manera en que los ciberatacantes se preparan, lanzan y ejecutan sus ataques.

Las tácticas son las categorías de las técnicas. Por tanto, son el qué los atacantes tratan de lograr. Las técnicas, individuales son el cómo logran sus objetivos. Dentro de la matriz las técnicas son las columnas, mientras que las técnicas son los distintos valores que poseen éstas. Es decir, es el comportamiento específico para alcanzar un objetivo dado. Esta relación se describe dentro de la matriz ATT&CK.

Initial Access	Execution	Persistence	Privilege Escalation	Defense Evasion	Credential Access	Discovery	Lateral Movement	Collection	Exfiltration	Command And Control
10 items	31 items	56 items	28 items	59 items	20 items	19 items	17 items	13 items	9 items	21 items
Drive-by Compromise	AppleScript	.bash_profile and .bashrc	Access Token Manipulation	Access Token Manipulation	Account Manipulation	Account Discovery	AppleScript	Audio Capture	Automated Exfiltration	Commonly Used Port
Exploit Public-Facing Application	CMSTP	Accessibility Features	Accessibility Features	Binary Padding	Bash History	Application Window Discovery	Application Deployment Software	Automated Collection	Data Compressed	Communication Through Removable Media
Hardware Additions	Command-Line Interface	AppCert DLLs	AppCert DLLs	BITS Jobs	Brute Force	Browser Bookmark Discovery	Distributed Component Object Model	Clipboard Data	Data Encrypted	Connection Proxy
Replication Through Removable Media	Control Panel Items	AppInit DLLs	AppInit DLLs	Bypass User Account Control	Credential Dumping	File and Directory Discovery	Exploitation of Remote Services	Data from Information Repositories	Data Transfer Size Limits	Custom Command and Control Protocol
Spearphishing Attachment	Dynamic Data Exchange	Application Shimming	Application Shimming	Clear Command History	Credentials in Files	Network Service Scanning	Logon Scripts	Data from Local System	Exfiltration Over Alternative Protocol	Custom Cryptographic Protocol
Spearphishing Link	Execution through API	Authentication Package	Bypass User Account Control	CMSTP	Credentials in Registry	Network Share Discovery	Pass the Hash	Data from Network Shared Drive	Exfiltration Over Command and Control Channel	Data Encoding
Spearphishing via Service	Execution through Module Load	BITS Jobs	DLL Search Order Hijacking	Code Signing	Exploitation for Credential Access	Password Policy Discovery	Pass the Ticket	Data from Removable Media	Exfiltration Over Other Network Medium	Data Obfuscation
Supply Chain Compromise	Exploitation for Client Execution	Bootkit	Dylib Hijacking	Component Firmware	Forced Authentication	Peripheral Device Discovery	Remote Desktop Protocol	Data Staged	Exfiltration Over Physical Medium	Domain Fronting
Trusted Relationship	Graphical User Interface	Browser Extensions	Exploitation for Privilege Escalation	Component Object Model Hijacking	Hooking	Permission Groups Discovery	Remote File Copy	Email Collection	Scheduled Transfer	Fallback Channels
Valid Accounts	InstallUtil	Change Default File Association	Extra Window Memory Injection	Control Panel Items	Input Capture	Process Discovery	Remote Services	Input Capture		Multi-hop Proxy
	Launchctl	Component Firmware	File System Permissions Weakness	DCShadow	Input Prompt	Query Registry	Replication Through Removable Media	Man in the Browser		Multi-Stage Channels
	Local Job Scheduling	Component Object Model Hijacking	Hooking	Deobfuscate/Decode Files or Information	Kerberoasting	Remote System Discovery	Shared Webroot	Screen Capture		Multiband Communication
	LSASS Driver	Create Account	Image File Execution Options Injection	Disabling Security Tools	Keychain	Security Software Discovery	SSH Hijacking	Video Capture		Multilayer Encryption
	Mshta	DLL Search Order Hijacking	Launch Daemon	DLL Search Order Hijacking	LLMNR/NBT-NS Poisoning	System Information Discovery	Taint Shared Content			Port Knocking
	PowerShell	Dylib Hijacking	New Service	DLL Side-Loading	Network Sniffing	System Network Configuration Discovery	Third-party Software			Remote Access Tools
	Regsvcs/Regasm	External Remote Services	Path Interception	Exploitation for Defense Evasion	Password Filter DLL	System Network Connections Discovery	Windows Admin Shares			Remote File Copy
	Regsvr32	File System Permissions Weakness	Plist Modification	Extra Window Memory Injection	Private Keys	System Owner/User Discovery	Windows Remote Management			Standard Application Layer Protocol
	Rundll32	Hidden Files and Directories	Port Monitors	File Deletion	Replication Through Removable Media	System Service Discovery				Standard Cryptographic Protocol
	Scheduled Task	Hooking	Process Injection	File System Logical Offsets	Securityd Memory					Standard Non-Application Layer Protocol
	Scripting	Hypervisor	Scheduled Task	Gatekeeper Bypass	Two-Factor Authentication Interception					Uncommonly Used Port
	Service Execution	Image File Execution Options Injection	Service Registry Permissions Weakness	Hidden Files and Directories						Web Service
	Signed Binary Proxy Execution	Kernel Modules and Extensions	Setuid and Setgid	Hidden Users						
	Signed Script Proxy Execution	Launch Agent		Hidden Window						
	Source			HISTCONTROL						
	Space after Filename			Image File Execution Options Injection						

Ilustración 8. Matriz ATT&CK

Por todo ello, con esta implementación se ha buscado cubrir cuatro elementos básicos:

- El comportamiento de los adversarios. Se centra en las tácticas y técnicas de los adversarios. En este caso el enfoque no se son las herramientas y el malware que usan los adversarios, sino en cómo interactúan con los sistemas durante una operación.

- Crear un modelo de ciclo de vida de las amenazas mejores a las existentes en la actualidad.

- La aplicabilidad a un entorno real. Se basa en incidentes reales. Por tanto, aplicable a entornos reales.

- Creación de una taxonomía común relacionada con las TTP.

Dentro de este framework podemos destacar dos matrices para las organizaciones: PRE-ATT&CK y ATT&CK para empresas. Ambas se combinan para cubrir todos los aspectos del CKC. La primera cubre las fases de reconocimiento y armado (Reconnaissance - Weaponization), mientras que la segunda cubre las restantes. Las tácticas de cada una son las siguientes:

<u>Tácticas PRE-ATT&CK.</u>

- Definición de prioridades.
- Selección de objetivos.
- Recopilación de información.
- Identificación de debilidades.
- Operaciones de seguridad adversas.
- Establecer y mantener la infraestructura.
- Desarrollo del personaje.
- Creación de funciones.
- Prueba de funciones.
- Implementación de funciones.

Tácticas de ATT&CK para empresas.

- Acceso inicial.
- Ejecución.
- Persistencia.
- Escalado de privilegios.
- Evasión de defensas.
- Acceso a credenciales.
- Detección.
- Movimiento lateral.
- Recopilación.
- Exfiltración.
- Comando y control.

Con MITRE ATT&CK se pueden conseguir algunos de los siguientes beneficios:

- Entendimiento cómo funciona el adversario. Cómo piensa, cuáles son sus necesidades, … todo esto ayuda a segurizar y defender de forma más eficiente las organizaciones. Permite por tanto, entender las fortalezas y debilidades actuales de la organización, y evolucionar hacia un estado más seguro.

- Permite la búsqueda eficiente de amenazas reales. Se puede incluso identificar al posible actor de un ataque sufrido.

- Intercambio de conocimientos entre los equipos de seguridad, referente a cómo puede actuar el adversario.

- Se trata de una base de conocimiento para la investigación. Esto facilita las labores de los equipos de analistas.

- Es una metodología de uso para los Red Team. Puede implementarse actividades de este tipo dentro de la organización para ayudar en su seguridad.

3.6. Sigma rules.

Una de las posibles aproximaciones a la hora de codificar las reglas que vamos a implementar es a través de las reglas Sigma, o también conocidas como Generic Signature Format for SIEM systems. Sigma es un formato de firma genérico y abierto que permite describir eventos relevantes. Dicho formato es flexible, fácil de escribir y aplicable a cualquier tipo de log. El objetivo con el que surgió este formato fue el unificar la forma en que se expresan las reglas para una fácil compartición entre la comunidad.

Gráficamente se describe de la siguiente forma:

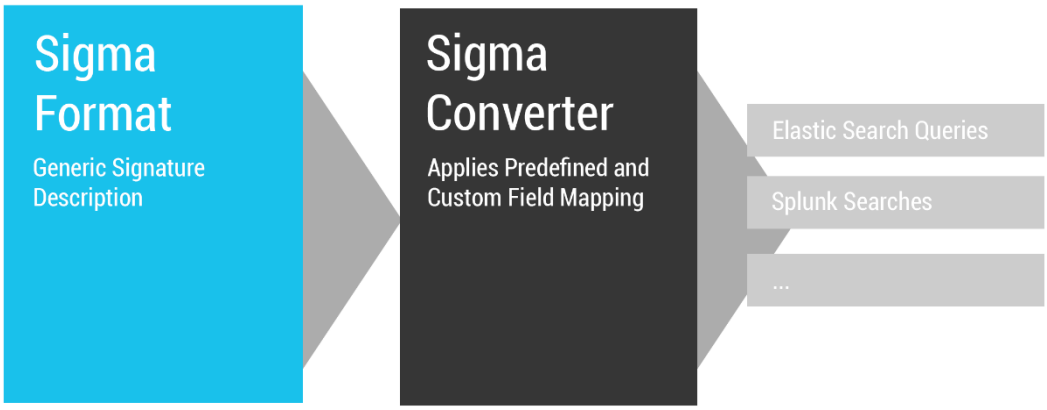

Ilustración 9. Sigma Rules

La gran ventaja que tiene escribir las alertas en Sigma es que se está haciendo en un lenguaje genérico, no dependiente del SIEM en con el cual se esté trabajando en este momento. Por tanto, las ventajas como se han visto son varias. El objetivo con el que nació este estándar abierto fue que se tratase de un mecanismo de detección en la que se definen, comparten y recopilan reglas para el beneficio de la comunidad.

Una vez que tenemos una regla escrita en lenguaje Sigma, podemos traducirla a un gran número de lenguajes SIEM: Splunk (consultas y dashboards), ElasticSearch Query Strings, ElasticSearch Query DSL, Kibana, Elastic X-Pack Watcher, Logpoint, Windows Defender Advanced Threat Protection (WDATP), Azure Sentinel / Azure Log Analytics, Sumologic, ArcSight, QRadar, Qualys, RSA NetWitness, PowerShell, Grep con compatibilidad de expresiones regulares de Perl y LimaCharlie.

Existe un conversor bidireccional de reglas Sigma a los diferentes SIEM en el siguiente enlace. También existe dentro del repositorio original del proyecto la posibilidad de instalar Sigmac para realizar las mismas funciones.

Si es necesario remarcar, que las reglas Sigma se ajustan perfectamente a consultas simples, pero no pueden implementar consultas complejas. Esto es muy importante de indicar dado que precisamente se pierde la potencialidad de Splunk si quisiéramos usar únicamente este formato. Por ejemplo, una de las grandes ventajas de Splunk es la posibilidad de realizar operaciones y transformaciones con los datos que se están usando dentro de las consultas. Estas acciones no pueden ser trasladadas al lenguaje de Sigma.

En este proyecto únicamente vamos a hacer uso de la obtención de inteligencia de la comunidad para pasar a implementar una serie de reglas dentro de Elastic. En todo caso, es un tema interesante a poder incluir en el futuro, principalmente pensando en posibles migraciones de reglas a otros SIEM.

3.7. Modo de conexión Elastic Stack con Splunk. HTTP Event Collector (HEC)

Como se ha comentado en múltiples ocasiones en el presente trabajo, lo que vamos a hacer es implementar un sistema de tipo Data Lake basado en Elastic Stack, y que las funciones SIEM las realizará el sistema Splunk Enterprise.

Para poder conectar ambos sistemas, vamos a hacer uso de una de las funcionalidades que nos permite Splunk a la hora de indexar eventos. Este sistema es el HTTP Event Collector (HEC). La forma de realizarlo a través de un HTTP request a través de un cliente que en este caso será la infraestructura del Elastic. Esta petición enviará un evento codificado en formato JSON.

Veamos cómo está configurado dentro de Splunk.

Para ver la configuración, seguimos los siguientes pasos dentro de Splunk:

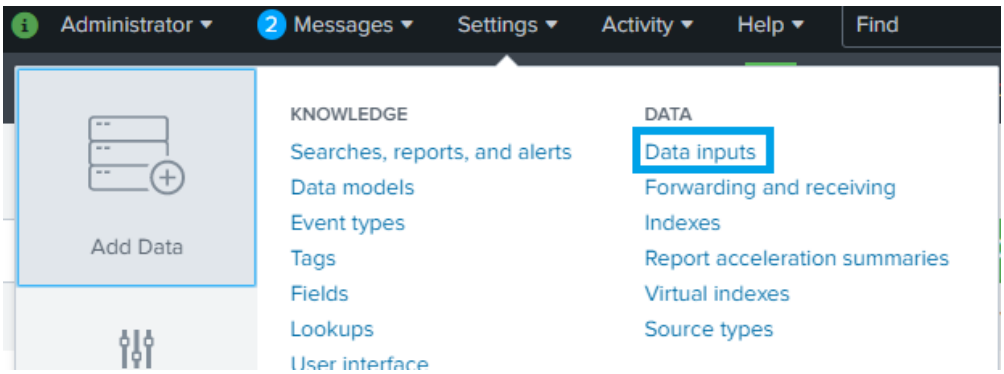

Ilustración 10. Settings > DATA > Data inputs

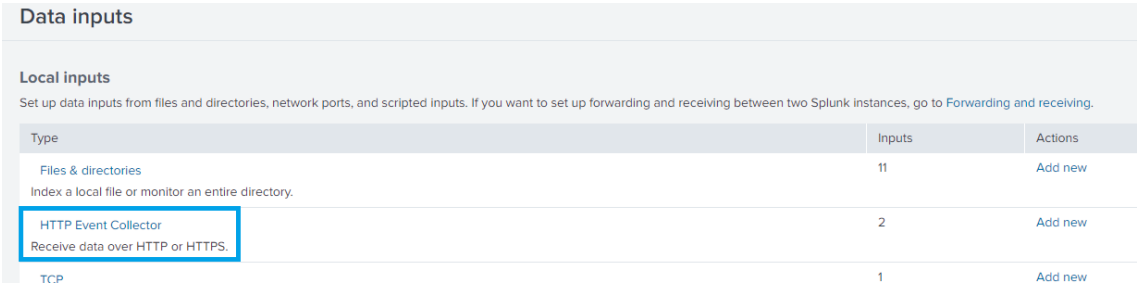

Ilustración 11. Type > HTTP Event Collector

Si accedemos dentro de las configuraciones globales, podemos ver qué valores están declarados:

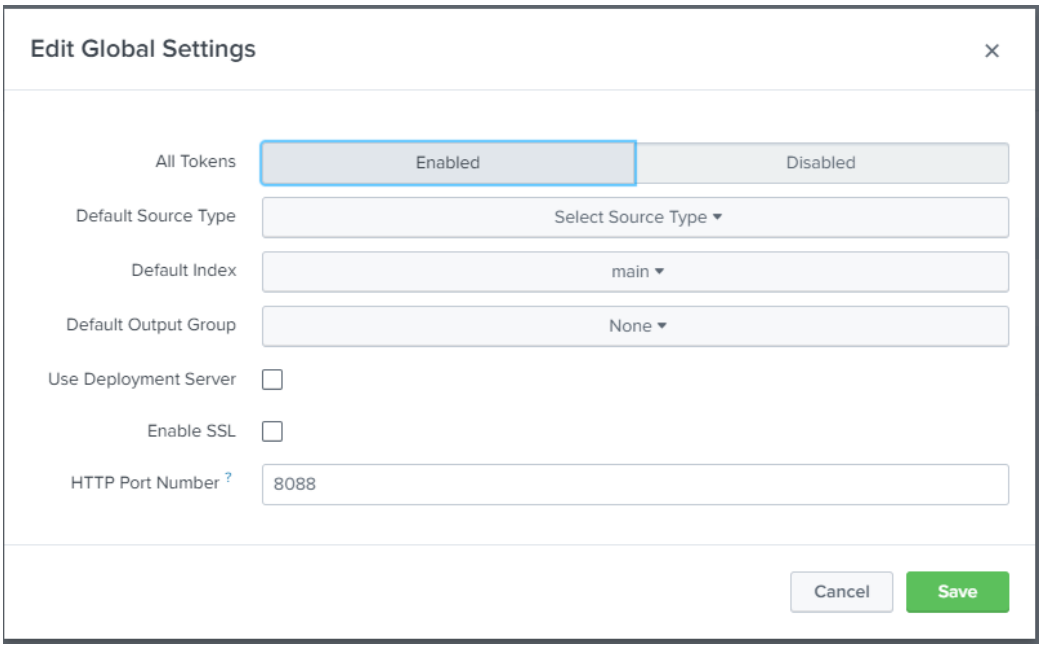

Ilustración 12. Edit Global Settings

Podemos destacar el número de puerto que se usa, y la posibilidad del envío sobre SSL. Para el presente trabajo no se va a utilizar esta función.

También podemos ver los dos canales creados para poder recibir eventos desde tanto el ElastAlert, como del Watcher.

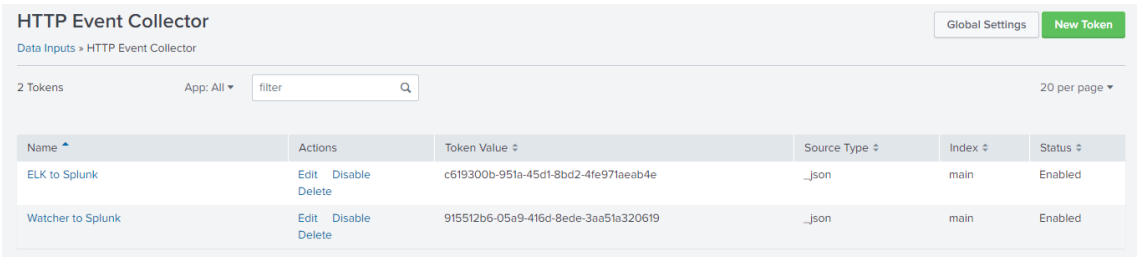

Ilustración 13. HTTP Event Collector

Se deben de definir unos simples parámetros. Lo principales son los siguientes:

- El índice donde se van a indexar los eventos recibidos. Por defecto en el presente proyecto se envían dentro del índice main. Se puede crear un índice particular para el indexado de estos eventos. Si deseásemos enviar los eventos a otro índice, bastaría con modificar el fichero inputs.conf dentro de la app donde se haya definido el HEC. En nuestro caso sería modificar dicho fichero dentro de la ruta $SPLUNK_HOME/etc/apps/search/local

- El tipo de eventos que le estamos enviando (sourcetype). En este caso, se trata de eventos de tipo JSON.

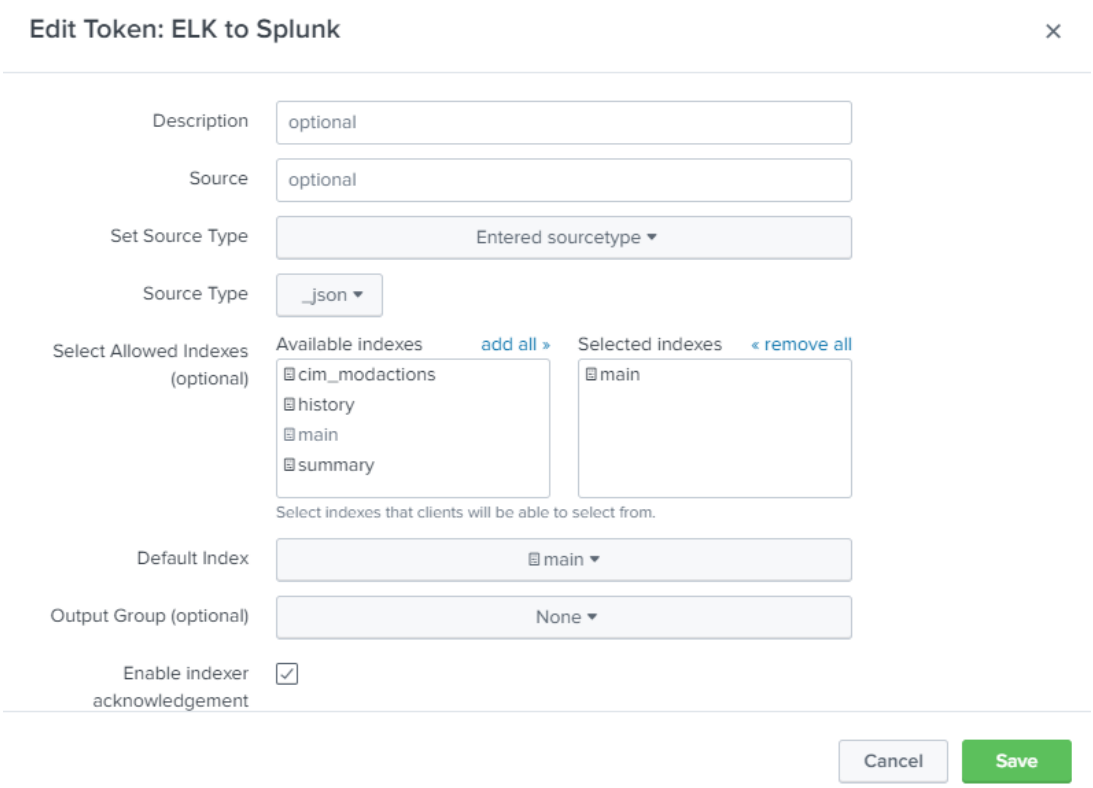

Ilustración 14. Edit Token: ELK to Splunk

Además, cuando se crea un canal de este tipo se nos facilitan dos elementos fundamentales para establecer la comunicación como son:

- Token. Que será añadirá dentro de la cabecera HTTP que le enviemos.
- Canal. Se deberá de definir por último un valor del canal. Para generar el GUID, podemos usar la siguiente página.

Esta configuración la podemos ver en la acción que se realiza cuando se dispara una alerta dentro de por ejemplo el ElastAlert.

```
alert: post
alert_subject: "NOMBRE_DE_LA_ALERTA"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD:8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_1: 'value_1'
   event_2: 'value_2'
   event_3: 'value_3'
   …
http_post_static_payload:
   static_event_1: 'value_1'
   static_event_2: 'value_2'
   …
http_post_headers:
X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
Accept: "*/*"
```

Del código anterior, hay que destacar como configuración para el éxito del envío los campos http_post_url, http_post_headers (X-Splunk-Request-Channel y Authorization).

La documentación oficial sobre HEC la podemos encontrar en el siguiente enlace.

3.8. Pasos a implementar.

A continuación, pasamos a enumerar los pasos que se deben de dar para poder implementar de forma práctica lo comentado anteriormente:

- Definición de las arquitecturas que se van a usar en el presente trabajo, tanto a nivel del Data Lake, como de Splunk Enterprise.

- Identificación de las fuentes origen de los eventos que se van a indexar dentro del Data Lake. Ya sean appliances de seguridad, como equipos finales.

- Una vez identificados las fuentes a enviar, vamos a determinar cuáles de todos sus eventos van a ser considerados como eventos notables. También se crearán las etiquetas relacionadas para cada EN.

- Normalización de los eventos anteriores, que serán enviados a través de HEC desde Elastic Stack a Splunk.

- Envío de los eventos desde el Data Lake hasta la arquitectura de SE, a través de la creación de una serie de alertas dentro de los correladores de Elastic.

- Asignación de las prioridades a las distintas fuentes que se van a monitorizar. En especial, se van a identificar aquellos elementos críticos para la organización (cuentas de administradores, servidores críticos, …). Se estudiará la ampliación de las etiquetas de los eventos notables.

- Análisis y creación de las distintas alertas dentro del SIEM.

- Generación de un Dashboard de gestión y monitorización de las alertas y los eventos notables.

- Creación de una serie de Playbooks para la respuesta ante incidentes basadas en las alertas creadas.

3.9. Arquitectura implementada – Elastic Stack.

La arquitectura implementada de Elastic está compuesta por los siguientes elementos.

- 1 cluster de dos máquinas de balanceadores.
- 8 nodos Logstash encargados de la recolección, procesamiento y redistribución de los datos. Están agrupados en parejas de dos balanceados. Cada pareja recibe unas fuentes de datos determinadas.
- 8 nodos ElasticSearch. Existen 6 nodos de tipo “data”, de estos 6 nodos, 3 son nodos de tipo “hot” y los otros 3 nodos son de tipo “cold”. Por último, y debido a cómo funciona Elastic, los otros 3 nodos que quedan son de tipo “master”.
- 1 Kibana. Existe una máquina con una instancia de Kibana instalada, que es la encargada de proporcionarnos la interfaz gráfica de usuario.

Es importante reseñar, que las fuentes en las que se recolectan eventos mediante agentes de Elastic (Winlogbeat, Packetbeat, …) envían sus datos directamente a los nodos “hot” de ElasticSearch. Las fuentes que no están normalizadas mediante formato JSON (envíos a través de por ejemplo syslog), van a pasar necesariamente por los Logstash.

Veamos a continuación, el diagrama de lo anterior. Se representan los elementos de Kibana. Las fuentes envían en primer lugar al balanceador, se omite por claridad en el diagrama.

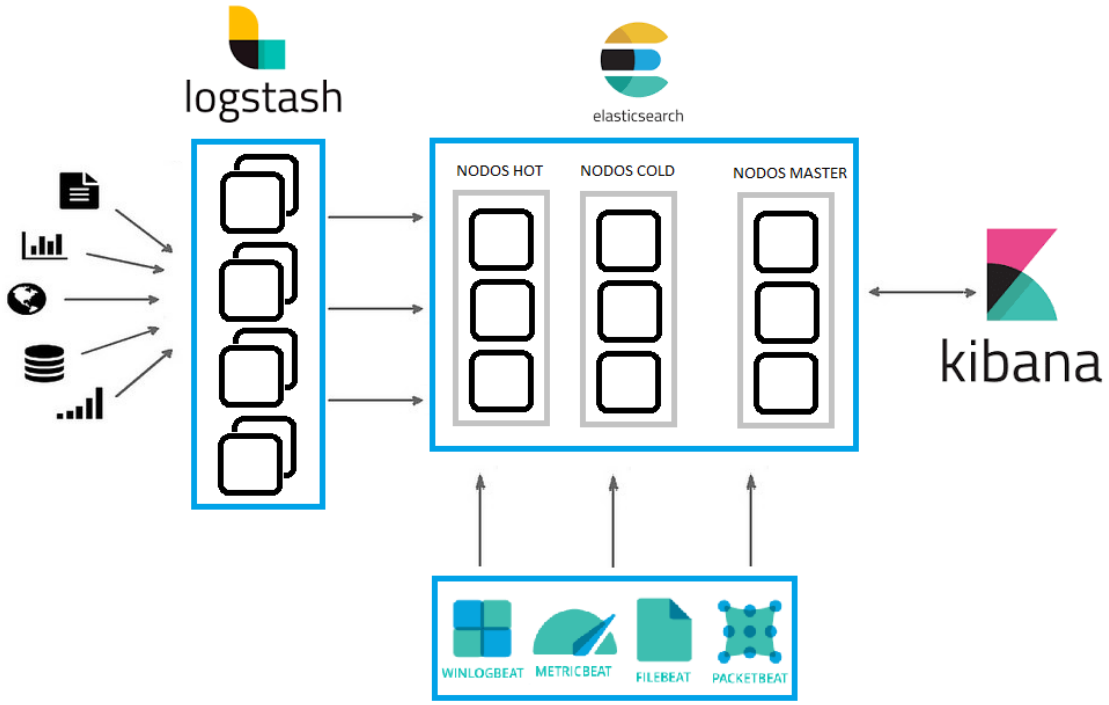

Ilustración 15. Arquitectura ELK

Como se va a ver a continuación, esta arquitectura es más grande que la existente dentro de Splunk. Esto es lógico, dado que la arquitectura de Elastic es la encargada de gestionar y almacenar todos los datos del Data Lake. Por tanto, estamos ante cientos de gigas de datos por día.

3.10. Arquitectura implementada - Splunk Enterprise.

La arquitectura con la que vamos a trabar está compuesta por los siguientes elementos:

- 1 Search Head.
- 1 cluster de indexadores compuesto por dos indexadores.
- 1 Heavy Forwarder.
- 5 Universal Forwarder.
- 1 Master Cluster Node.
- 1 License Server.
- 1 Deployment Server.

Además, y a través del HEC, tenemos la conexión con el Data Lake.

A continuación, podemos ver esto gráficamente:

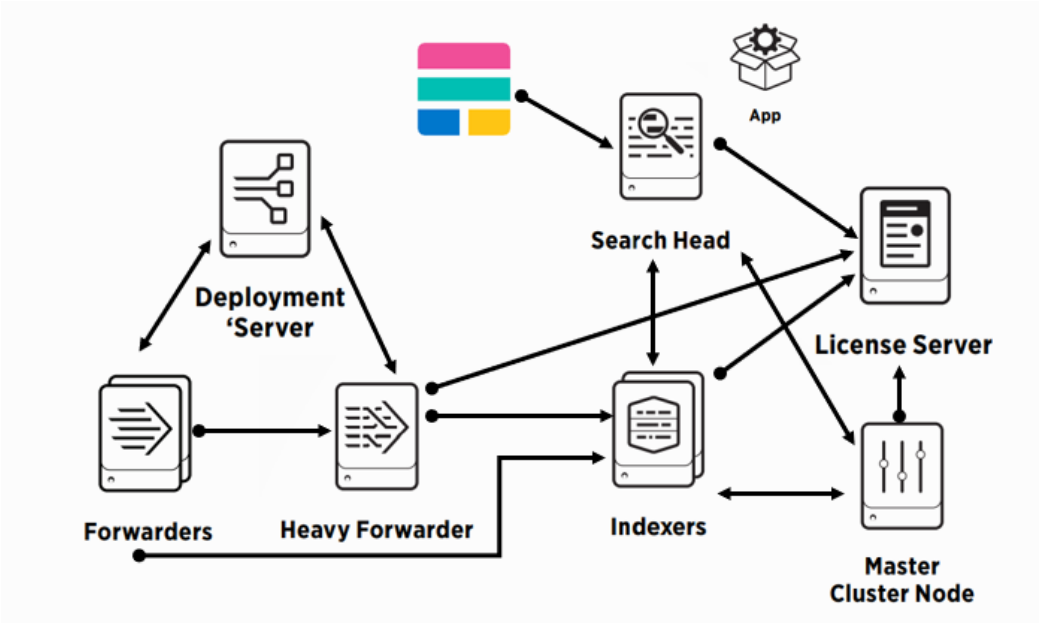

Ilustración 16. Arquitectura Splunk

No se va a profundizar demasiado en la configuración de la arquitectura distribuida dado que no es este el propósito del trabajo. Anteriormente se ha descrito algunas nociones sobre Splunk, en este trabajo nos vamos a centrar en otros elementos más enfocados al campo del analista de seguridad, que de la parte relacionada con la integración de soluciones de seguridad.

Mediante esta arquitectura vamos a desarrollar la solución SIEM del presente trabajo. A continuación, se va a hacer referencia a dos aspectos a destacar:

- Dado que se va a poseer una licencia comercial de Splunk, aunque esta sea mínima, lo que se va a hacer es aprovecharla para indexar directamente algunas fuentes de datos. Van a ser algunas fuentes que por distintas razones, se ha decidido que no sean indexadas dentro del

Data Lake. En todo caso, no afecta en ningún caso en el concepto global del proyecto. Al contrario, todo queda integrado de forma armónica.

- Dada la sinergia entre los fabricantes de aplicativos de seguridad de la industria y Splunk, vamos a hacer uso de las APP desarrolladas para la integración vía API de elementos que van a enriquecer los datos contenidos dentro del SIEM.

3.11. Identificación de las fuentes Data Lake.

En este apartado vamos a ver una serie de fuentes que están enviando eventos dentro del Data Lake, y qué eventos podemos sacar de éstas para poder ser enviados al SIEM.

A modo teórico vamos a describir las fuentes principales que se suelen recolectar dentro de una organización.

- Dispositivos de red: Routers, switches, bridges, puntos de acceso Wireless, modems, hubs.
- Servidores: Web, proxy, mail, FTP, bases de datos, controladores de dominio, servidor LDAP, servidor DNS, servidores de correo, …
- Dispositivos de seguridad: IDS/IPS, firewalls, EDR, EPP, dispositivos de filtrado de contenido, aplicaciones de detección de intrusos, Honeypots, Concentradores VPN, DLP, UTM, …
- Aplicaciones: Aplicaciones propietarias, aplicaciones de terceros y comerciales, IoT, …

Como puede verse, realmente casi cualquier elementos de la organización puede ser monitorizado. Sin embargo, y dependiendo de las necesidades de la empresa u organización a monitorizar, primaran unos sobre otros. En todo caso, siempre existirán una serie de fuentes que deberán ser monitorizados. Estamos hablando de los eventos provenientes de los appliance de seguridad, servidores críticos, o equipos finales de usuarios críticos.

Ahora sí, vamos a ver sobre qué fuentes vamos a realizar el análisis para generar los eventos notables. Dado que es un ejercicio teórico, se va a limitar el número de fuentes a implementar.

- Eventos de auditoría y Sysmon de Windows.
- Eventos del EDR Cisco AMP.
- Eventos del Firewall de nueva generación Palo Alto.
- Eventos de Cisco ESA.

Es muy importante el indicar que lo que vamos a hacer aquí es únicamente una aproximación a lo que se debería de hacer en caso de implementar esta solución en un sistema real. Por tanto, tanto las fuentes que vamos a tratar en el presente trabajo, como dentro de éstas los tipos de eventos procesados, van a ser una guía de cómo hacer las cosas, más que un ejemplo real y completo de cómo quedaría dentro de un sistema en producción. Simplemente, se deberá seguir desarrollando el trabajo hasta su conclusión.

3.12. Eventos notables. Creación y envío.

3.12.1. Eventos sistema Windows.

Eventos de Windows. Eventos canal de Seguridad, Aplicación y Sistema.

En este apartado vamos a describir qué eventos de Windows se han seleccionado para enviar desde Elastic a Splunk. Veremos la forma en que se recolectan a nivel de sistema y como se mandan al Data Lake. Y una vez dentro de éste, como los enviamos a Splunk.

Los eventos de auditoría de los sistemas Windows, y en especial el canal de Seguridad, son una de las fuentes más recomendables de monitorizar. Existen una gran bibliografía al respecto generadas tanto por analistas de seguridad como por empresas del sector que analizan qué eventos son los más interesantes dentro del canal de Seguridad. También hay que tener en cuenta, que dependiendo del sistema donde se estén generando dichos eventos, su identificador puede cambiar.

Pero empecemos por el principio. Estos eventos pertenecen a una serie de políticas de auditoría que están implementadas dentro de los sistemas Windows. Para poder acceder a estos, debemos hacer uso de la herramienta denominada Event Viewer (Visor de eventos). Será por tanto la política de auditoría la que determinará los eventos del sistema Windows que serán almacenados dentro del registro de seguridad. Windows posee un total de nueve categorías de políticas de auditoría, junto a 50 subcategorías. De esta forma se proporciona una gran granularidad. No es objetivo del presente trabajo entrar en detalle de todo esto, así que se facilita toda la información sobre la temática en el siguiente enlace. El enlace anterior redirige a una web referencia a nivel mundial sobre los eventos de seguridad de Windows.

Por tanto, estamos ante una serie de eventos que se generan dentro de los sistemas Windows y que tienen relevancia a nivel de seguridad. Mediante las políticas de auditoría se determinan qué eventos serán almacenados para su posible posterior análisis. Cada evento puede ser auditado en caso de éxito, de fallo, o ambas. Y por supuesto, puede descartase el proceso de auditado.

Indicar también, que los eventos que se auditan dependerán del sistema sobre el que nos estemos refiriendo. Habrá mayor posibilidad de auditoría si estamos en un servidor Windows con funciones de DC, que sobre un Workstation normal. Además, el identificador cambia dependiendo si son eventos de sistemas antiguos (win2000, XP y Win2003) o los actuales.

Dentro del canal de seguridad de Windows podemos encontrar cientos de eventos distintos. Podemos imaginar la cantidad de cientos de miles de logs que se podrán generar dentro de un red con equipos Windows cada día. Esto nos va a obligar a tener que realizar un filtrado grande en base a los identificadores de los eventos, y a la cantidad de estos que se generan al día.

A continuación, podemos ver una lista de eventos de Windows que suelen ser recomendados el monitorizar:

Event ID	Event summary
4624	An account was successfully logged on
4625	An account failed to log on
4648	A logon was attempted usina explicit credentials
4719	System audit policy was changed
4964	Special groups have been assigned to a new logon
1102	The audit log was cleared
4720	A user account was created
4722	A user account was enabled
4723	An attempt was made to change an account's password
4725	A user account was disabled
4728	A user was added to a privileged global group
4732	A user was added to a privileged local group
4735	A security-enabled local group was changed
4756	A user was added to a privileged universal group
4738	A user account was changed
4740	A user account was locked out
4767	A user account was unlocked
4737	A privileged global group was modified
4755	A privileged universal group was modified
4772	A Kerberos authentication ticket request failed
4777	The domain controller failed to validate the credentials of an account.
4782	Password hash an account was accessed
4616	System time was changed
4657	A registry value was changed
4697	An attempt was made to install a service
4698	A scheduled task was created
4699	A scheduled task was deleted
4700	A scheduled task was enabled
4701	A scheduled task was disabled
4702	A scheduled task was updated
4946	A rule was added to the Windows

	Firewall exception list
4947	A rule was modified in the Windows Firewall exception list
4950	A setting was changed in Windows Firewall
4954	Group Policy settings for Windows Firewall has changed
5025	The Windows Firewall service has been stopped
5031	Windows Firewall blocked an application from accepting incoming traffic
5152	The Windows Filtering Platform blocked a packet
5153	A more restrictive Windows Filtering Platform filter has blocked a packet
5155	Windows Filtering Platform blocked an application or service from listening on a port
5157	Windows Filtering Platform blocked a connection
5447	A Windows Filtering Platform filter was changed

Tabla 6. Event ID de Windows a implementar como EN

Tenemos que entender que cualquier eventos de seguridad puede ser de interés a la hora de analizar un incidente de seguridad. Por tanto, lo que estamos haciendo aquí es un filtrado selectivo para la generación de los eventos notables y alertas. Sin embargo, el resto de eventos pueden ser consultados en un posterior análisis forense de logs, o incluso si no todos los eventos son enviados a través del canal de suscripción, podemos desplegar un agente (tanto a nivel de Elastic Stack a través de su Winlogbeat, como a través de Splunk y su UF) para una recolección selectiva, o por último, nos podemos conectar a la máquina en concreto y a través del visor de eventos, analizar todos los eventos generados (una forma que exista un usuario dentro del equipo con permisos de lectura de los event logs de las máquinas de dominio).

Si analizamos la lista anterior, podemos ver que falta un evento fundamental y que no se ha incluido. Para este evento vamos a realizar un análisis especial, dado que es un evento fundamental. El evento en concreto es el event id 4624 – Successful User Account Login. Este evento monitoriza el login exitoso de un usuario dentro del sistema. Se ha excluido inicialmente de los eventos a monitorizar debido al gran volumen de eventos que se general al día de estos eventos.

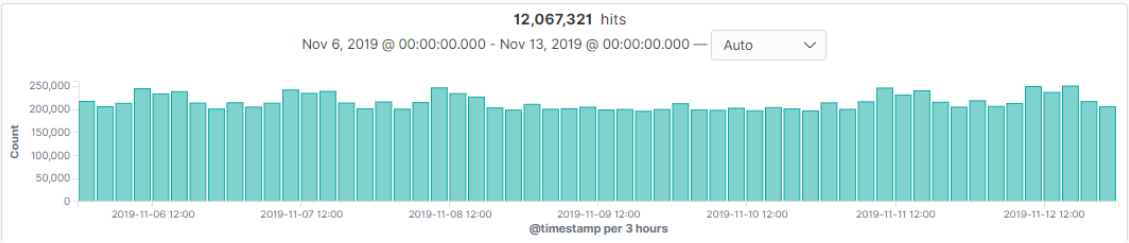

Ilustración 17. Eventos totales 4624

Una posible solución es monitorizar únicamente los eventos de algún tipo especial como pueden ser los login a través de una conexión remota (RemoteInteractive).

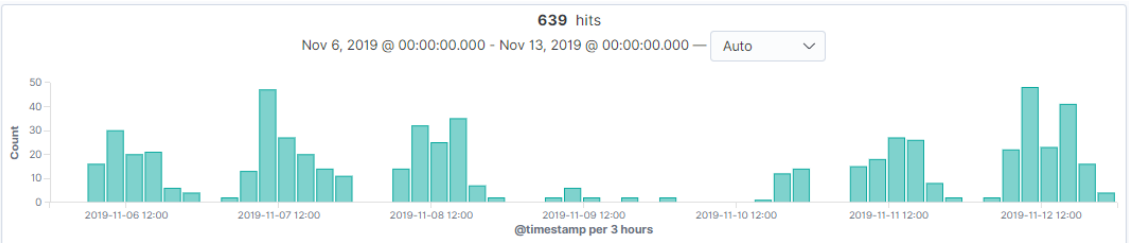

Ilustración 18. Eventos con LogonType 10

En base a los resultados de la consulta anterior, podríamos perfectamente determinar el enviar este tipo de eventos a Splunk.

Nombre de la alerta	Descripción de la alerta
EN_Windows_4624_10	Alerta que se dispara cuando se detecta la generación de un evento de tipo 4624 dentro de un sistema Windows, con la particularidad de que el logon Type debe de ser 10.

Tabla 7. Alerta de Windows del evento 4624 Logon Type 10

Veamos a continuación el código de la alerta dentro del correlador ElastAlert.

```
#Alert EN_Windows_4624_10
_host: IP_ELASTIC
run_every:
  minutes: 1
bufferime:
  minutes: 2
es_port: 9200
_source_enabled: true
name: EN_Windows_4624_10
description: Seguridad
type: frequency
query_key: "@timestamp"
index: logstash-winlogbeat-*
num_events: 1
include:
["event_data.IpAddress","event_data.LogonType","event_data.TargetUserN
ame","event_data.WorkstationName","event_id","log_name","computer_name
"]
realert:
  hours: 0
timeframe:
```

```
  minutes: 1
filter:
        - query:
                query_string:
                   query: "event_id:4624 AND event_data.LogonType: 10"
alert: post
alert_subject: "EN_Windows_4624_10"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_type: '@timestamp'
   src_ip: 'event_data.IpAddress'
   Logon_Type: 'event_data.LogonType'
   user: 'event_data.TargetUserName'
   src_host: 'event_data.WorkstationName'
   signature_id: 'event_id'
   Log_Name: 'log_name'
   src: 'computer_name'
   prioridad: '60'
http_post_static_payload:
   alerta: 'EN_Windows_4624_10'
   tags: ‘authentication,windows,security,os,remote,AoO’
http_post_headers:
X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
Accept: "*/*"
```

En otro ordenes de cosas, no solo debemos monitorizar eventos del canal de Seguridad. Dentro de los eventos de auditoría podemos encontrar de otros tipos como pueden ser los de Aplicación, Configuración o Sistema. Dentro de estas categorías podemos encontrar algunos eventos interesantes de monitorizar como son los siguientes:

Event ID	Event log	Event Summary
1000	Application	Application Error
1002	Application	Application Hang
1001	Application	WER (Windows Error Reporting)
1001	System	BSOD (Microsoft-Windows-WER-SystemErrorReporting)
104	System	Event Log was Cleared
102	Security	Audit Log was Cleared
4719	System	System audit policy was changed
1125	System	Internal Error
1127	System	Generic Internal Error
1129	System	Group Policy Application Failed due to Connectivity

Tabla 8. Otros eventos interesantes de Windows

A continuación, vamos a ver el número total de eventos estimados que se van a filtrar y enviar desde Elastic a Splunk.

Para poder orientarnos sobre el número de eventos que vamos a tener que tratar, se ha realizado el siguiente ejercicio. Vamos a ver cuántos eventos de Windows hay en una ventana de tiempo de 7 días consecutivos.

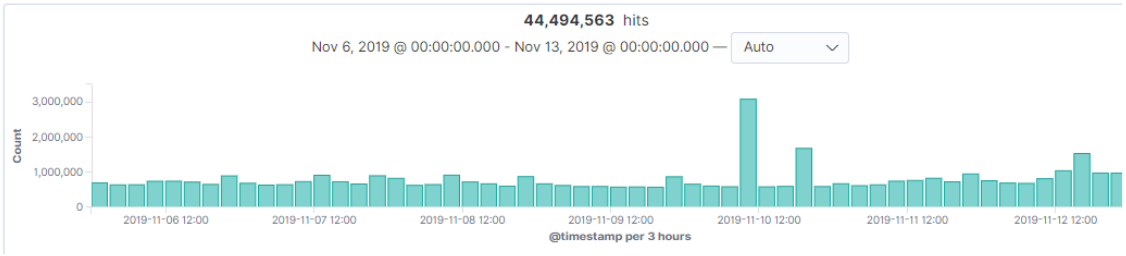

Ilustración 19. Total eventos Windows 7 días

Como se puede ver, son casi 45 millones de eventos. Veamos de qué canales estamos recibiendo eventos:

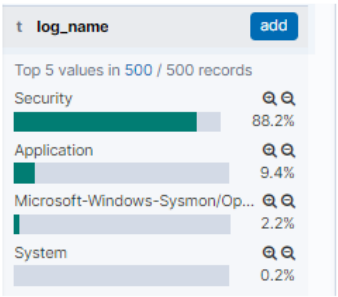

Ilustración 20. Canales Windows

Por tanto un poco más de 88% son del canal de seguridad. Ahora vamos a ver cuántos eventos tenemos si excluimos los dos eventos más volumétricos: 4624 y 4634.

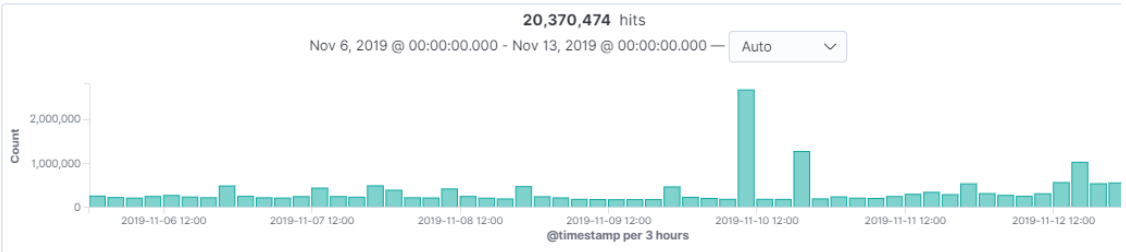

Ilustración 21. Exclusión eventos 4624 y 4634

Excluyendo únicamente estos dos eventos, se reduce a más de la mitad los eventos indexados. Ahora vamos a filtrar el resultado, mostrando únicamente los eventos que vamos a reenviar a Splunk:

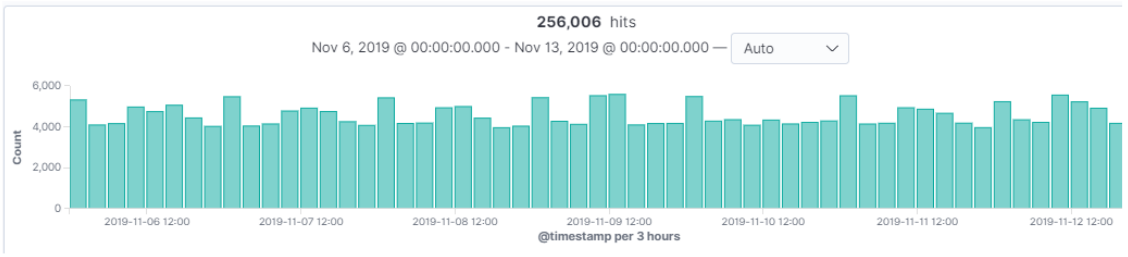

Ilustración 22. Número de eventos totales a reenviar

Son un poco más de 250K eventos. Este número es mucho más manejable que el inicial.

Eventos de Windows. Sysmon.

Para conseguir una mejor monitorización de los sistemas Windows, se decidió instalar la aplicación de Sysmon dentro de los sistemas Windows monitorizados.

Sysmon es una herramienta desarrollada por Mark Russinovich y Thomas Garnier, que es gratuita, propiedad de Microsoft. Dicha herramienta pertenece a un grupo grande de herramientas de sistemas, denominado Sysinternals.

Sysmon es un servicio del sistema de Windows, a la vez que un controlador de dispositivo, que se encarga de supervisar y registrar la actividad del sistema. Genera por tanto una serie de eventos, que son almacenados dentro del registro de sucesos de Windows. Sysmon proporciona por ejemplo, información detallada sobre la creación de procesos, conexiones de red, y cambios en el tiempo de creación de archivos entre otros muchos.

Por tanto, se encarga de monitorizar toda la actividad del sistema donde está instalado. Es muy importante indicar que el detalle de la monitorización dependerá de la configuración que le indiquemos.

Aunque no es el objetivo del presente trabajo el profundizar en la instalación y gestión de Sysmon, sí que se van a indicar desde dónde se puede descargar Sysmon de forma oficial, y un par de páginas [1] y [2], donde nos guiaran además de la propia oficial en su instalación y manejo.

El tratamiento de los eventos de Sysmon se puede abordar de tres formas distintas. Una primera aproximación podría ser el envío de todos los eventos que se generan de este tipo en los distintos sistemas. En este caso estaríamos hablando del siguiente número de eventos para la ventana de tiempo anterior:

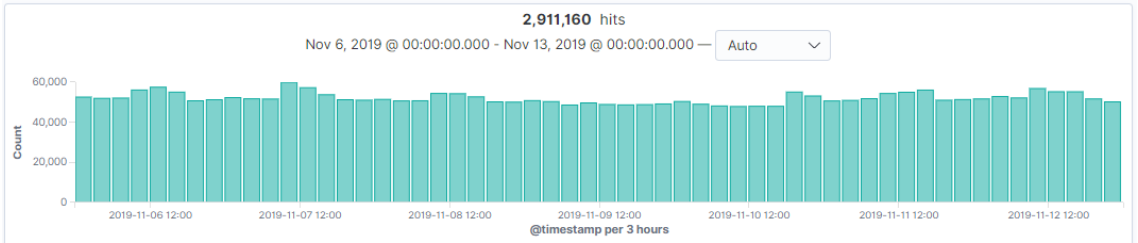

Ilustración 23. Número eventos totales Sysmon

Como se puede ver, su número es bastante alto como para ser enviados como eventos notables. Veamos pues, otras opciones.

Otra segunda forma sería, dada la complejidad del envío de todos estos eventos de la forma en que la hemos planteado, la creación de una serie de alertas para detectar aquellos eventos realmente importantes. Bajo esta premisa, nos podríamos basar en las reglas ya existentes dentro del proyecto Sigma relacionadas con Sysmon. Si se eligiese este enfoque, sería tan fácil como transformar las reglas Sigma existentes dentro del repositorio de GitHub del proyecto, al lenguaje Watcher. Lo podríamos hacer de forma casi automática desde la siguiente aplicación online.

Por último, la tercera forma sería el hacer uso de la herramienta OSSEC-Wazuh dado que posee sus propias reglas de monitorización de Sysmon.

Para el presente trabajo, se ha determinado dado que se tiene desplegado, el hacer uso de las reglas existentes dentro de Wazuh sobre Sysmon. Para ello, más adelante veremos un ejemplo de cómo hacer esto. Se ha elegido este método, dado que se considera que es el más fácil y preciso a nivel de correlación y de implementación. Esto es así, dado que Wazuh no deja de ser ya un SIEM con sus propias alertas correladas. Sí indicar, que en el caso de no poseer la posibilidad de tener instalado Wazuh, la segunda solución también sería una buena aproximación.

En todo caso, se indicó que existía la posibilidad de indexar ciertos datos directamente dentro de Splunk (en base a la licencia mínima necesaria de contratación). Y dado que estos eventos suelen ser muy interesantes a la hora de analizar y correlar, siempre se podría optar por enviarlos directamente a Splunk a través de un Universal Forwarder instalado dentro del servidor WEF.

Proceso de recolección y envío de los eventos de Windows.

No se va a entrar en detalle dado que no es el objetivo del presente trabajo, pero sí se va a indicar el modo de recolección llevado a cabo. Los logs son recolectados a través de WEF (Windows Event Forwarding), gracias al uso de WinRM. Este proceso se basa en el envío de los eventos de Windows a través de la red a un servidor central WEF. El proceso se basa en la creación de un canal de suscripción dentro del servidor. Los distintos clientes envían los logs al servidor, siendo filtrados por lo declarado en dicho canal. En la suscripción se define qué eventos serán reenviados al servidor WEF. Es por tanto un filtro de los eventos a recibir.

Se ha decidido enviar al Data Lake todos los eventos que se generan tanto en los canales Security, como los del canal de Application, System y los generados a raíz de la instalación de Sysmon (Microsoft-Windows-Sysmon/Operational).

Una vez que tenemos los eventos dentro del servidor WEF, mediante un agente de tipo Winlogbeat, los enviamos a la infraestructura de Elastic. De esta

forma, tendremos indexados dentro del Data Lake todos los eventos sin filtro de las máquinas monitorizadas a través de WEF.

Por último indicar, que se poseerá un servidor WEF por cada uno de los dominios existentes a nivel de Directorio Activo.

Envío y normalización de los eventos notables desde Elastic a Splunk.

A continuación, vamos a mostrar un par de ejemplos de cómo debemos proceder para enviar los eventos desde Elastic a Splunk. Se describirá la selección de los campos a enviar, la normalización, la asignación de las distintas etiquetas y la implementación de la alerta que se encargará del envío de los eventos a Splunk.

Evento 4625.

En el primer ejemplo, se ha decidido el tratamiento del evento 4625. Lo primero que debemos hacer es ver cómo tenemos almacenado la información de dicho evento dentro de Elastic.

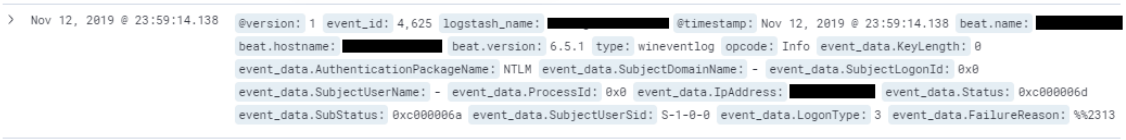

Ilustración 24. Evento ejemplo 4625

Nunca debemos de olvidar que estamos trabajando con un fichero de tipo JSON:

Ilustración 25. Evento 4625 en formato JSON

El paso más importante que vamos a tener que dar aquí, es determinar qué campos son realmente importantes a la hora de enviar a Splunk. Podría parecer que lo ideal es enviar todo el log, pero esto en este modelo no es lo mejor. Primero, por un problema de espacio de almacenamiento. Más concretamente por un problema con la licencia de Splunk. Por otro, porque vamos a considerar el SIEM que estamos desarrollando como un correlador de eventos a analizar. Pero cuando realmente queramos ver todos los eventos y toda la información, lo que haremos será consultar los logs almacenados dentro del Data Lake. Por todo ello, debemos analizar evento a evento, determinando qué campos son realmente los interesantes.

Para el ejemplo en el que estamos, vamos a analizar qué campos son los interesantes. Estamos ante un evento de login fallido, por lo tanto, debemos obtener los siguientes campos clave para realizar una correlación correcta:

- Tenemos el apartado *Account For Which Logon Failed*. Esto es, el usuario que ha intentado logarse fallidamente.
- Luego tenemos el tipo de logon que se ha usado.
- El origen de la máquina donde se realiza la petición.
- Y por último, y más importante tenemos el *Failure Information*. Esto es un código (Status y Sub Status Codes) que nos va a indicar el motivo del fallo.

Por todo lo anterior, nos debemos quedar con los siguientes campos dentro del log del Kibana:

- event_data.IpAddress. Máquina origen donde se produce el fallo de logon.
- event_data.LogonType. Tipo de logon que se ha realizado (ver Logon Type del evento 4624).
- event_data.Status y event_data.SubStatus. Códigos de error.
- event_data.TargetUserName. Nombre de la cuenta que ha intentado el logon.
- event_data.WorkstationName. Nombre de la máquina donde se ha producido el error.
- event_id. Número del evento. Genérico.
- log_name. Canal del evento de Windows. Genérico.
- computer_name. DC donde se ha logado el evento. Genérico.
- @timestamp. Hora de generación del evento. Genérico.
- Tags. Etiquetas generadas por nosotros.

Como se puede ver, se ha reducido mucho el número de campos del evento (11/44) y su tamaño.

Otro paso clave del proceso es dado que vamos a enviar los eventos a Splunk normalizados, comprobar la normalización CIM de Windows y adaptar los campos extraídos a dicha normalización.

Para ello nos vamos a la propia documentación de Splunk referente al CIM, y buscamos los campos equivalentes. Como sabemos, debemos de dirigirnos a la siguiente página. Ciertamente, puede llegar a ser complicado el ir identificando campo a campo su equivalente dentro de los Dataset definidos.

Se crea la siguiente tabla de equivalencias:

Campo Elastic	Campo Splunk	Dataset name
event_data.IpAddress	src_ip	Authentication
event_data.LogonType	NO CIM	
event_data.Status	Status	Endpoint
event_data.SubStatus	NO CIM	
event_data.TargetUserName	user	Authentication

event_data.WorkstationName	src_host	Authentication
event_id	signature_id	Authentication
log_name	NO CIM	
computer_name	src	Authentication

Tabla 9. Tabla de equivalencias

Como se puede ver, no es fácil el identificar el campo normalizado al que corresponde cuando estamos manejando eventos tan complejos como son los de auditoría de un sistema. Incluso, podemos encontrarnos campos que no pertenecen a CIM (recordar, que CIM no es en este sentido demasiado estricto). Por ello, debemos incluso tomar la decisión de determinar qué nombre de campo se va a considerar para el estándar fuera de CIM.

Otra forma totalmente válida es usar la TA de parseo oficial de Windows que podemos encontrar dentro de Splunkbase. La forma de trabajar de esta forma será una vez descargado y descomprimido, nos dirigimos al fichero Splunk_TA_windows\default\transforms.conf y Splunk_TA_windows\default\props.conf para analizar la forma en cómo Splunk parsea los eventos de Windows para estar conforme a CIM. Veamos un ejemplo.

Se puede enfocar de dos formas, una más compleja, que es analizar dicho fichero buscando la expresión regular que parsea el campo deseado, o se puede si se tiene instalado el Add-on en Splunk, indexar un evento en bruto de tipo deseado, y ver cómo es parseado por dicha TA.

El evento 4625 lo podemos meter dentro de Splunk de dos formas distintas a nivel de raw. En texto claro, o como XML. En todo caso, Splunk es capaz de parsear eventos estructurados. Esto hará que cuando analicemos el evento, veamos que para un mismo valor existen múltiples campos, que además, no suelen cumplir con CIM. Por ello, una vez listados los campos, se deberá comprobar cuál es el que corresponde si existe, con CIM.

A continuación veamos la tabla de equivalencias:

Campo Elastic	WinEventLog	XmlWinEventLog
event_data.IpAddress	src_ip Source_Network_Address	src_ip IpAddress
event_data.LogonType	Logon_Type	LogonType Logon_Type
event_data.Status	Error_Code Status ta_windows_status	Error_Code Status ta_windows_status
event_data.SubStatus	Sub_Status	SubStatus Sub_Status
event_data.TargetUserName	Account_Name member_dn user	TargetUserName Target_User_Name user

event_data.WorkstationName	src Workstation_Name src_nt_host	src Source_Workstation WorkstationName src_nt_host
event_id	EventCode signatura_id	EventCode EventID signatura_id
log_name	LogName	Channel
computer_name	dest ComputerName dest_nt_host dvc	src Source_Workstation WorkstationName src_nt_host

Tabla 10. Tabla equivalencias Event ID 4625

Como se ha comentado, podemos ver para cada campo, varios equivalentes. Algunos son generados en base a la extracción automática que realiza Splunk, pero otros son generados por la aplicación directa de la TA de Windows. En este caso, debemos determinar qué campos son CIM y cuáles no. Existe para ayudarnos dentro de Splunkbase, una APP que nos ayuda a determinar si estamos o no cumpliendo CIM.

En base al análisis anterior, la tabla de equivalencias quedará de la siguiente forma:

Campo Elastic	**Campo Splunk**
event_data.IpAddress	src_ip
event_data.LogonType	Logon_Type
event_data.Status	Status
event_data.SubStatus	Sub_Status
event_data.TargetUserName	User
event_data.WorkstationName	src_host
event_id	signature_id
log_name	Log_Name
computer_name	src

Tabla 11. Tabla equivalencia final

Se ha determinado por tanto en referencia a los campos que no son CIM, el aplicar un nombre de campo en el que se separan las palabras por el carácter '_', y cada una de estas palabras se escriben en minúsculas, salvo la primera letra que se declara como mayúscula.

Únicamente faltaría definir las etiquetas para este evento. Se han determinado asignarle a este evento las siguientes etiquetas:

- authentication
- failure

- windows
- security
- error
- os

Estas etiquetas se han basado en las existentes dentro de la propia TA de Windows. No siempre será así, habrá casos en que se determinará que es necesario añadir otro tipo de etiquetas ad-hoc.

Una vez que tenemos claro los campos de la alerta dentro del ElastAlert, y los campos que se van a enviar a Splunk, pasamos a ver la alerta que se ha creado. Vamos a mantener un estándar a la hora de crear estas alertas. La nomenclatura para los eventos de Windows será de la siguiente forma: EN_Source_ID.yaml Para el caso en concreto del evento 4625, el nombre del fichero será por tanto el siguiente: EN_Windows_4625.yaml

Nombre de la alerta	Descripción de la alerta
EN_Windows_4625	Alerta que se dispara cuando se detecta la generación de un evento de tipo 4625 dentro de un sistema Windows.

Tabla 12. Tabla de alertas Windows

A continuación podemos ver el código de la alerta creada:

```
#Alert EN_Windows_4625
_host: IP_ELASTIC
run_every:
  minutes: 1
bufferime:
  minutes: 2
es_port: 9200
_source_enabled: true
name: EN_Windows_4625
description: Seguridad
type: frequency
query_key: "@timestamp"
index: logstash-winlogbeat-*
num_events: 1
include:
["event_data.IpAddress","event_data.LogonType","event_data.Status","ev
ent_data.SubStatus","event_data.TargetUserName","event_data.Workstatio
nName","event_id","log_name","computer_name"]
realert:
  hours: 0
timeframe:
  minutes: 1
filter:
        - query:
                query_string:
                        query: "event_id:4625"
alert: post
alert_subject: "EN_Windows_4625"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
```

```
    event_type: '@timestamp'
    src_ip: 'event_data.IpAddress'
    Logon_Type: 'event_data.LogonType'
    Status: 'event_data.Status'
    Sub_Status: 'event_data.SubStatus'
    user: 'event_data.TargetUserName'
    src_host: 'event_data.WorkstationName'
    signature_id: 'event_id'
    Log_Name: 'log_name'
    src: 'computer_name'
    prioridad: '20'
http_post_static_payload:
    alerta: 'EN_Windows_4625'
    tags:
"authentication,failure,windows,security,error,os,exploitation"
http_post_headers:
X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
Accept: "*/*"
```

A continuación podemos ver el evento indexado dentro de Splunk:

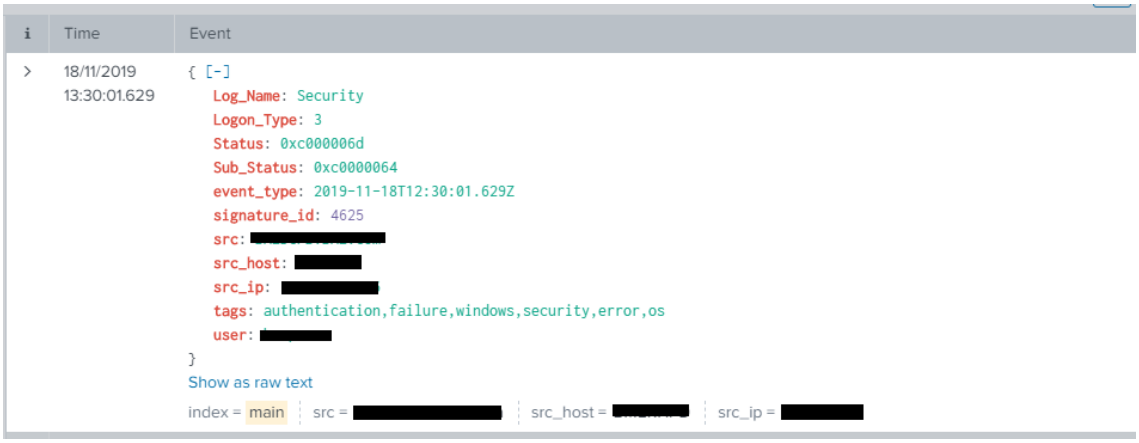

Ilustración 26. Evento 4625 indexado dentro de Splunk

Se puede ver también, cómo es el evento _raw que se recibe.

Ilustración 27. Evento 4625 en formato _raw

Como es la primera alerta que se implementa detalladamente, vamos a explicarla por bloques para una mejor comprensión.

Podríamos diferenciar dos grupos principales. El primero sería la configuración de la alerta, mientras que el segundo es la configuración de la acción de la alerta cuando se dispara ésta.

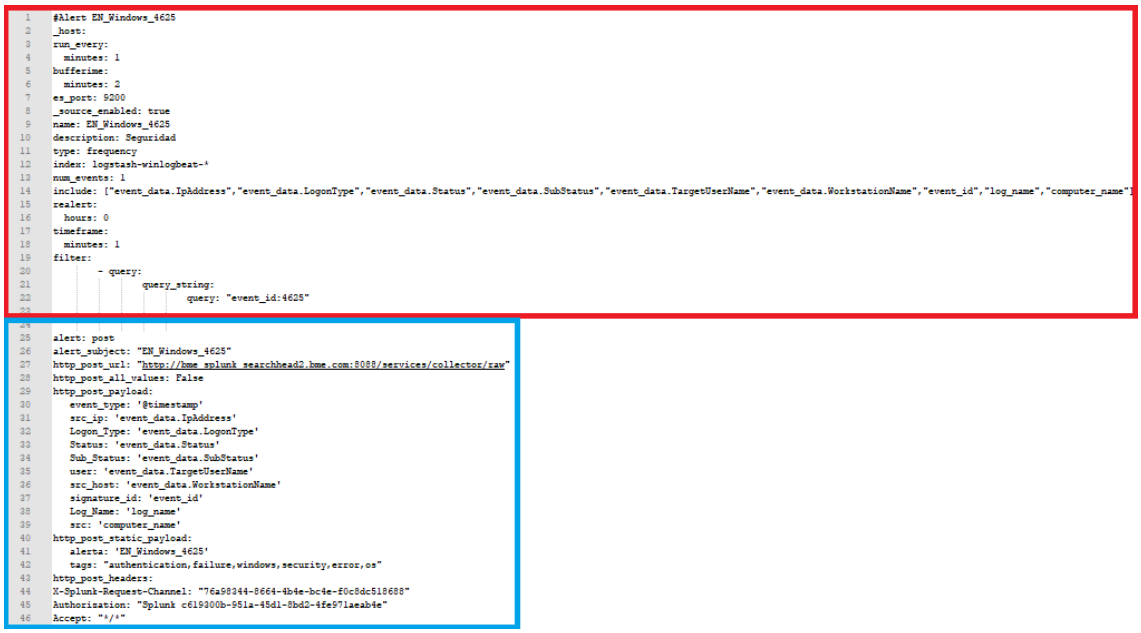

Ilustración 28. Alerta ElastAlert EN_Windows_4625

La configuración de la alerta a su vez podría separarse en dos partes. La primera sería la configuración del tipo de alerta que se va a implementar, junto con las distintas marcas de tiempo. Mientras que la segunda sería la parte donde se define el filtro de búsqueda. De forma unificada podemos destacar los siguientes campos:

- La alerta está planificada para ejecutarse cada minuto (función run_every).
- Se indica el índice donde se va a realizar la consulta (función index).
- Se define el tipo dentro de ElastAlert (función type). En este caso es de tipo frequency. La alerta se dispara cuando supera el umbral definido en num_events.
- La alerta va a comprobar su filtro dentro de la ventana de tiempo que se le indica dentro del argumento timeframe.
- Por último, debemos de destacar la función filter. Es aquí, donde se define el filtro que se va a aplicar como consulta a los eventos dentro de Elastic. Siendo por tanto, la principal función de correlación.

Por último, la configuración de qué acción debe de realizar la alerta cuando se dispara es la siguiente:

- Se define el tipo de respuesta que se va a aplicar. En este caso mediante el siguiente parámetro. alert: post
- Mediante la función http_post_url, se define la ruta a dónde se va a enviar el post.
- El payload del post se define tanto en http_post_static_payload y http_post_payload.
- Por último, se declara el método http_post_headers, donde se definen los parámetros para permitir la conexión entre ambos sistemas.

Acabamos de ver la mayor problemática que posee este proyecto. Y no es otra que el tener que ir seleccionando los eventos que queremos enviar, los campos que son útiles dentro de estos, y la normalización de dichos campos. Este trabajo es el más laborioso de realizar.

Se podría plantear el realizar la normalización de los eventos en la parte de Splunk, pero también sería un proceso manual, dado que no existe una TA para el parseo de eventos provenientes de un Winlogbeat. Además, hay que recordar que Elastic indexa los eventos ya parseados en formato JSON, por tanto no es viable el enviar a Splunk un log en formato raw.

Otra acción que podría tomarse y que ahorraría tiempo, es enviar el log en bruto (sin filtro de campos) a Splunk. El problema que tendríamos si adoptásemos este método es que dado que tenemos una licencia limitada dentro de Splunk, es posible que la superásemos. El filtrado de campos nos facilita tanto la posibilidad de indexar una mayor cantidad de eventos dentro de Splunk, como un mejor entendimiento de los eventos.

Evento Sysmon Suspicious Process – explorer.exe.

Nombre de la alerta	Descripción de la alerta
EN_Windows-Wazuh_001	Alerta que se dispara cuando se detecta la creación del proceso explorer.exe, con un proceso padre extraño. Alerta basada en la propia correlación de Wazuh.
EN_Windows_explorer	Misma alerta que la primera, se genera en base a la correlación de un evento de Sysmon.

Tabla 13. Tabla de aletas Sysmon

Como se ha comentado un poco más arriba, se ha elegido para la creación de eventos de Sysmon, el usar los propios eventos que genera Wazuh. Para poder entender cómo correla este SIEM, vamos a analizar las partes de las reglas que hacen que se disparen estos eventos. Para ello vamos a analizar la rule relacionada con Sysmon.

Las reglas de Wazuh son una serie de ficheros de tipo XML que definen qué eventos y con qué “level” se van a indexar y notificar. El nivel de notificación también se define dentro de la configuración del Wazuh.

```
<group name="windows,sysmon,">
  <rule id="61600" level="0">
    <if_sid>60004</if_sid>
    <field name="win.system.severityValue">^INFORMATION$</field>
    <description>Windows Sysmon informational event</description>
    <options>no_full_log</options>
  </rule>
  <rule id="61603" level="0">
    <if_sid>61600</if_sid>
    <field name="win.system.eventID">^1$</field>
       <description>Sysmon    -    Event    1:    Process    creation
$(win.eventdata.description)</description>
    <options>no_full_log</options>
    <group>sysmon_event1,</group>
```

```
  </rule>
…
<group name="windows,sysmon,sysmon_process-anomalies,">
…
  <rule id="61640" level="12">
    <if_group>sysmon_event1</if_group>
    <field name="win.eventdata.image">\\explorer.exe</field>
         <description>Sysmon      -      Suspicious      Process      -
explorer.exe</description>

<group>pci_dss_10.6.1,pci_dss_11.4,gdpr_IV_35.7.d,hipaa_164.312.b,nist
_800_53_AU.6,nist_800_53_SI.4,</group>
  </rule>
  <rule id="61641" level="0">
    <if_sid>61640</if_sid>
    <field name="win.eventdata.parentImage">userinit.exe</field>
       <description>Sysmon    -    Legitimate    Parent    Image    -
explorer.exe</description>
  </rule>
…
```

Veamos pues, el código de la alerta creado dentro de ElastAlert:

```
# Alert EN_Windows-Wazuh_001
es_host: IP_ELASTIC
es_port: 9200
name: EN_Windows-Wazuh_001
description: Seguridad
type: frequency
run_every:
  minutes: 1
buffer_time:
  minutes: 2
realert:
  hours: 0
index: wazuh-alerts-3.x-*
query_key: "@timestamp"
include:
["data.win.eventdata.utcTime","agent.name","data.win.eventdata.user","
data.win.eventdata.parentImage","data.win.eventdata.parentCommandLine"
,"data.win.eventdata.image"]
num_events: 1
timeframe:
  minutes: 1
filter:
- bool:
        must:
                - query_string:
                        query: "rule.id: 61640"
alert: post
alert_subject: "EN_Windows-Wazuh_001"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_time: '@timestamp'
   src: 'agent.name'
   user: 'data.win.eventdata.user'
   Parent_Command_Line: 'data.win.eventdata.parentCommandLine'
   Image: 'data.win.eventdata.image'
   Command_Line: 'data.win.eventdata.commandLine'
```

```
   prioridad: '40'
http_post_static_payload:
   alerta: 'EN_Windows-Wazuh_001'
   tags: "suspicious,sysmon,security,os,installation"
http_post_headers:
   X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
   Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
   Accept: "*/*"
```

Y a continuación, vemos la indexación dentro de Splunk resultado del disparo de la alerta.

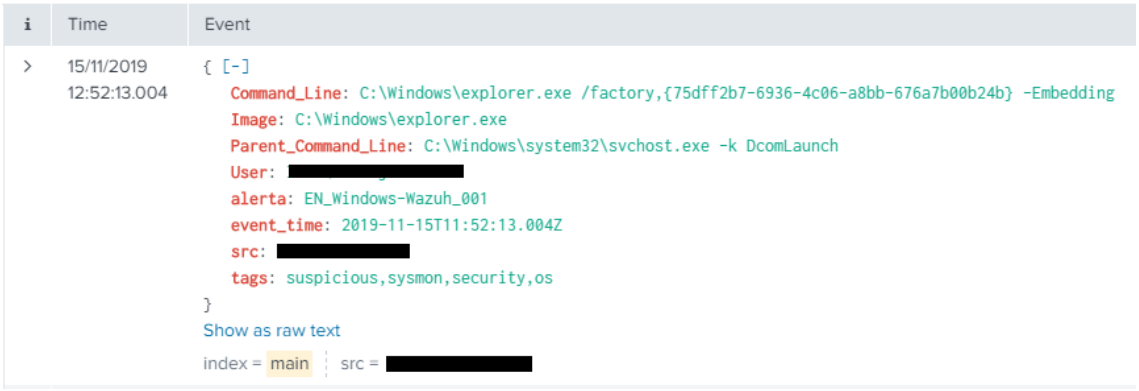

Ilustración 29. EN_Windows-Wazuh_001 indexado dentro de Splunk

En el caso de que no pudiésemos implementar Wazuh dentro de nuestra arquitectura, lo que sí podríamos hacer es crear las reglas de forma manual. Las rules de Wazuh son consultables por ejemplo desde aquí.

Vamos a ver a continuación cómo quedaría la implementación de la regla anterior correlada dentro del ElastAlert.

Antes de crear el código, vamos a ver qué resultados podemos esperar de esta regla. Lo vemos en base a una consulta dentro del Kibana.

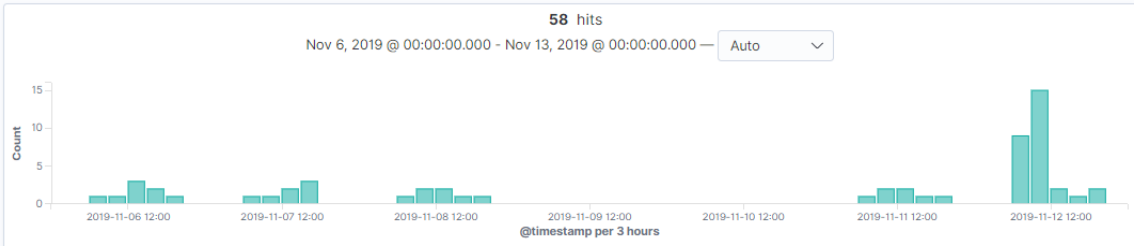

Ilustración 30. Eventos Windows Explorer

Como se puede ver, en la ventana de tiempo con la que hemos estado trabajando, se generarían un total de 58 eventos.

```
# Alert EN_Windows_explorer
es_host: IP_ELASTIC
es_port: 9200
name: EN_Windows_explorer
description: Seguridad
type: frequency
run_every:
```

```
  minutes: 1
buffer_time:
  minutes: 2
realert:
  hours: 0
index: logstash-winlogbeat-*
query_key: ["@timestamp","computer_name"]
include:
["computer_name","user.name","user.domain","event_data.User","event_da
ta.CommandLine","event_data.ParentCommandLine","event_data.Image"]
num_events: 1
timeframe:
  minutes: 1
filter:
- bool:
        must:
                - query_string:
                            query:    "log_name:    \"Microsoft-Windows-
Sysmon/Operational\"    AND    event_id:    1    AND    event_data.Image:
\"C:\\\\Windows\\\\explorer.exe\""
alert: post
alert_subject: "EN_Windows_explorer"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_time: '@timestamp'
   src: 'computer_name'
   user: 'event_data.User'
   Parent_Command_Line: 'event_data.ParentCommandLine'
   Image: 'event_data.Image'
   Command_Line: 'event_data.CommandLine'
   prioridad: '40'
http_post_static_payload:
   alerta: 'EN_Windows_explorer'
   tags: "suspicious,sysmon,security,os,installation"
http_post_headers:
   X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
   Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
   Accept: "*/*"
```

A continuación, podemos ver un evento indexado dentro de Splunk.

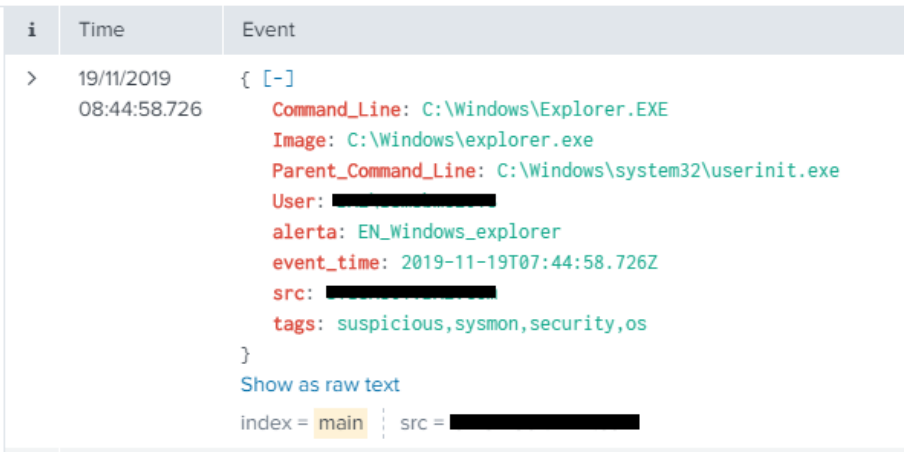

Ilustración 31. Evento indexado dentro de Splunk

Como se ha podido ver, se han seleccionado una serie de eventos dentro de los sistemas Windows que se pueden considerar como importantes a la hora de monitorizar. El trabajo que quedaría es ir haciendo evento a evento, lo realizado anteriormente.

También indicar que en todo caso, en cualquier caso podemos crear una alerta que envíe otros eventos que consideremos necesarios. Con estos eventos pasaremos posteriormente a realizar correlaciones más complejas.

3.12.2. Eventos Cisco AMP.

En esta sección vamos a indicar la forma en que vamos a tratar los eventos de generados por un EDR (Endpoint Detection and Response). En concreto, nos vamos a centrar en el EDR de Cisco denominado AMP (Advanced Malware Protection).

No se va a profundizar en qué consiste un EDR, ni en las características de Cisco AMP, dado que no son objetivo de éste, habiendo además múltiple documentación online al respecto. Sí vamos a comentar que esta solución estará desplegada dentro de todos los equipos finales de usuario de la organización.

La forma de integrarlo con el Data Lake va a ser a través la API de la aplicación. De esta forma, los eventos de detección relevantes serán indexados dentro del Elastic. Es interesante indicar que si bien se reciben los eventos más importantes dentro de un sistema detectado por el AMP, el resto de las actividades que son monitorizadas y logeadas dentro de la base de datos del AMP (se almacena dentro de un Elasticsearch), no son accesibles a través de la API. Esto será muy importante en el momento en que se quiera realizar una análisis de un posible incidente, dado que no solo tendremos que recurrir a la información almacenada dentro del Data Lake, sino que también podremos y deberemos analizar lo almacenado dentro de los logs del AMP.

Vamos a crear una alerta que recoja cualquier evento generado dentro del EDR, dado que se consideran todos relevantes. Posteriormente lo que haremos, será realizar correlaciones con dichos eventos en la parte SIEM.

Nombre de la alerta	Descripción de la alerta
EN_CiscoAMP	Alerta genérica que envía a Splunk cualquier evento generado dentro del Cisco AMP.

Tabla 14. Tabla de alertas Cisco AMP

```
# Alert EN_CiscoAMP
es_host: IP_ELASTIC
es_port: 9200
name: EN_CiscoAMP
description: Seguridad
type: frequency
run_every:
  minutes: 1
buffer_time:
  minutes: 2
```

```
realert:
  hours: 0
index: logstash-ciscoamp*
query_key: ["@timestamp","hostname"]
include:
["hostname","event_type","internal_ip","file_name","file_disposition",
"fingerprint","event.detection"]
num_events: 1
timeframe:
  minutes: 1
filter:
- bool:
        must:
                - query_string:
                        query: "hostname: *"
alert: post
alert_subject: "EN_CiscoAMP"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_time: '@timestamp'
   dest: 'hostname'
   category: 'event_type'
   dest_ip: 'internal_ip'
   signature: 'event.detection'
   file_name: 'file_name'
   file_hash: 'file_fingerprint'
   action: 'file_disposition'
   prioridad: '40'
http_post_static_payload:
   alerta: 'EN_CiscoAMP'
   tags: "malware,amp,attack,alert,exploitation"
http_post_headers:
   X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
   Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
   Accept: "*/*"
```

A continuación, podemos ver un evento indexado dentro de Splunk:

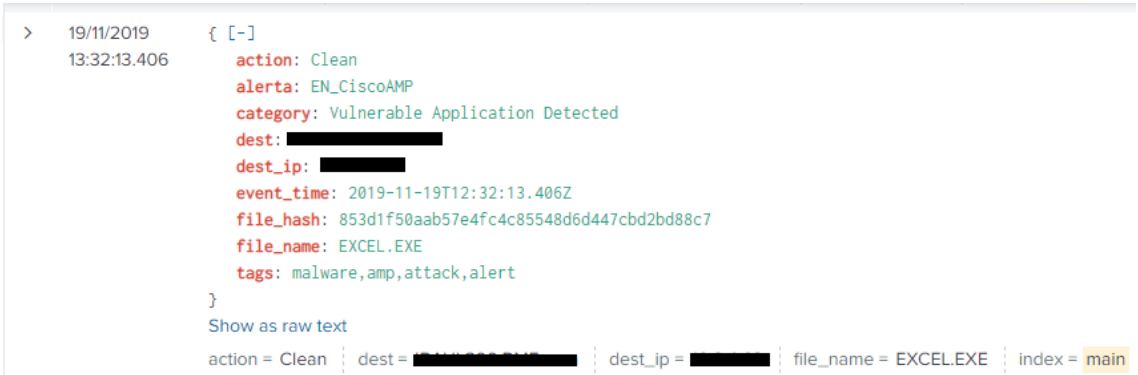

Ilustración 32. Evento EN_CiscoAMP indexado dentro de Splunk

Para la normalización de los eventos a recibir dentro de Splunk, se ha utilizado la propia TA que se puede descargar de Splunkbase.

3.12.3. Eventos Firewall - Palo Alto.

Otra de las fuentes que vamos a monitorizar es un FW de nueva generación. No es objetivo del presente trabajo el entrar en detalle en la configuración de un dispositivo de este tipo. Sin embargo, sí vamos a indicar que, obviamente, los eventos que tengamos a disposición para trabajar con ellos dependerán de dicha configuración (políticas de acceso, acciones a llevar a cabo, verbosidad del evento, …). Esto hará que nuestros eventos notables y posteriores alertas dependan fuertemente de dicha configuración. En todo caso, y con el objetivo de que el presente trabajo pueda aplicarse en un ámbito genérico, cuando entremos dentro de la configuración de las alertas dentro del SIEM, implementaremos alertas genéricas con el objetivo descrito anteriormente.

Para el presente proyecto, se van a desarrollar dos alertas de ejemplo, basadas en las capacidades actuales de un FW de nueva generación como es un Palo Alto.

Nombre de la alerta	Descripción de la alerta
EN_Firewall_001	Alerta que detecta eventos entrantes (Internet – Interno) que sean detectadas por el appliance como un posible ataque. Se envía al SIEM toda la información referente a la conexión entrante. Alerta dependiente de la configuración que se tenga aplicada al Firewall.
EN_Firewall_002	Alerta que detecta eventos entrantes desde una geolocalización extraña. En este caso, se detectan eventos provenientes de China. Origen considerando como de monitorización relevante. Alerta aplicable en el caso de que se monitorice el origen geográfico de la conexión entrante.

Tabla 15. Tabla alertas Firewall

```
# Alert EN_Firewall_001
es_host: IP_ELASTIC
es_port: 9200
name: EN_Firewall_001
description: Seguridad
type: frequency
run_every:
  minutes: 1
buffer_time:
  minutes: 2
realert:
  hours: 0
index: filebeat-panos-*
query_key: ["@timestamp","client.ip","destination.ip"]
include:
["client.ip","client.port","destination.ip","destination.port","event.
action","event.outcome","panw.panos.threat.name","panw.panos.ruleset"]
num_events: 1
timeframe:
  minutes: 1
filter:
- bool:
```

```
        must:
                - query_string:
                        query: "event.category: security_threat AND
NOT event.outcome: alert AND NOT event.action: url_filtering AND
panw.panos.ruleset: Internet-Interna"
alert: post
alert_subject: "EN_Firewall_001"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_time: '@timestamp'
   src_ip: 'client.ip'
   src_port: 'client.port'
   dest_ip: 'destination.ip'
   dest_port: 'destination.port'
   event_action: 'event.action'
   vendor_action: 'event.outcome'
   signature: 'panw.panos.threat.name'
   rule: 'panw.panos.ruleset'
   prioridad: '40'
http_post_static_payload:
   alerta: 'EN_Firewall_001'
   tags: "attack,network,firewall,incoming,exploitation"
http_post_headers:
   X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
   Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
   Accept: "*/*"
```

Evento correctamente indexado dentro de Splunk:

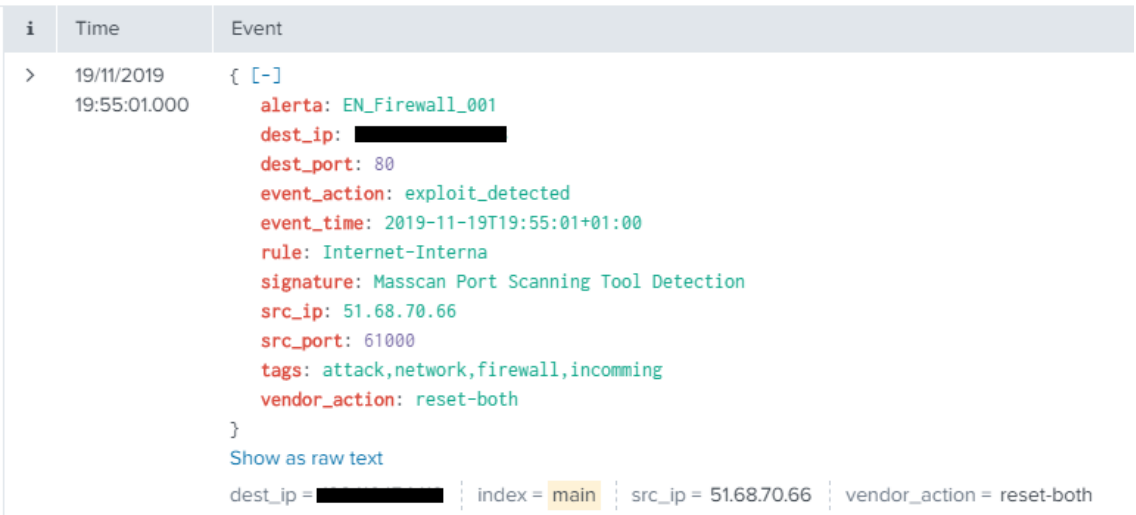

Ilustración 33. Evento EN_Firewall_001 indexado en Splunk

```
# Alert EN_Firewall_002
es_host: IP_ELASTIC
es_port: 9200
name: EN_Firewall_002
description: Seguridad
type: frequency
run_every:
  minutes: 1
buffer_time:
  minutes: 2
```

```
realert:
  hours: 0
index: filebeat-panos-*
query_key: ["@timestamp","client.ip","destination.ip"]
include:
["client.ip","client.port","destination.ip","destination.port","event.
action","event.outcome","panw.panos.threat.name","panw.panos.ruleset",
"source.geo.country_iso_code"]
num_events: 1
timeframe:
  minutes: 1
filter:
- bool:
        must:
               - query_string:
                            query:     "event.module:     panw     AND
source.geo.country_iso_code: CN"
alert: post
alert_subject: "EN_Firewall_002"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_time: '@timestamp'
   src_ip: 'client.ip'
   src_port: 'client.port'
   dest_ip: 'destination.ip'
   dest_port: 'destination.port'
   event_action: 'event.action'
   vendor_action: 'event.outcome'
   signature: 'panw.panos.threat.name'
   rule: 'panw.panos.ruleset'
   prioridad: '35'
http_post_static_payload:
   alerta: 'EN_Firewall_002'
   tags: "attack,network,firewall,incoming,recon,C2"
http_post_headers:
   X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
   Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
   Accept: "*/*"
```

i	Time	Event
>	20/11/2019 12:01:39.000	{ [-] alerta: EN_Firewall_002 dest_ip: [redacted] dest_port: 23 event_action: flow_terminated event_time: 2019-11-20T12:01:39+01:00 rule: GEO Protect signature: null src_ip: 106.13.167.159 src_port: 36895 tags: attack,network,firewall,China vendor_action: allow } Show as raw text dest_ip = [redacted] index = main src_ip = 106.13.167.159 vendor_action = allow

Ilustración 34. Evento EN_Firewall_002 indexado en Splunk

En cuanto a alertas de este tipo, la librería de posibles eventos notables puede ser muy extensa. Se pueden generar alertas en base a un único evento, como de forma volumétrica en búsqueda de anomalías. En la parte referente a las alertas del SIEM desarrollaremos más profundamente este concepto.

3.12.4. Eventos Cisco ESA.

Por último, vamos a ver la creación de eventos notables relacionados con una appliance de seguridad de correo electrónico. Cuando se implementan alertas basadas en tecnologías particulares, dependemos fuertemente de la configuración y capacidades de dicho software (como se ha visto en el apartado anterior referente a los FW). Se han implementado un total de dos alertas de forma detallada basadas en las capacidades de dicho appliance. Supondremos dado que no se puede asegurar, que todos los appliance de seguridad relacionados con la seguridad de las pasarelas de correos, poseen capacidades similares.

Nos hemos basado en dos capacidades intrínsecas del Cisco ESA, a saber, que posea un AV incorporado, y que posea la capacidad de en base a la reputación, detectar URLs sospechosas.

Nombre de la alerta	Descripción de la alerta
EN_ESA_001	Alerta que detecta un correo entrante que ha sido bloqueado por la acción del AV.
EN_ESA_002	Alerta que detecta un correo entrante, con una URL que posee una reputación mala.

Tabla 16. Tabla de alertas Cisco ESA

```
# Alert EN_ESA_001
es_host: IP_ELASTIC
es_port: 9200
name: EN_ESA_001
description: Seguridad
type: frequency
run_every:
  minutes: 1
buffer_time:
  minutes: 2
realert:
  hours: 0
index: logstash-ironport*
query_key: ["@timestamp","iron_from","ion_to"]
include:
["iron_from","iron_domain_fromheader","iron_to","iron_subject_full","i
ron_attachment","iron_tagging","iron_av","iron_engine"]
num_events: 1
timeframe:
  minutes: 1
filter:
- bool:
        must:
                - query_string:
                        query: "logger: finished AND iron_av: \"using
Sophos VIRAL\""
alert: post
```

```
alert_subject: "EN_ESA_001"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_time: '@timestamp'
   src_user: 'iron_from'
   from_header: 'iron_domain_fromheader'
   recipient: 'iron_to'
   subject: 'iron_subject_full'
   attachment: 'iron_attachment'
   antivirus_status: 'iron_av'
   spam_verdict: 'iron_engine'
   prioridad: '20'
http_post_static_payload:
   alerta: 'EN_ESA_001'
   tags: "attack,email,malware,quarantine,delivery"
http_post_headers:
   X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
   Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
   Accept: "*/*"
```

Viendo un ejemplo de evento indexado dentro de Splunk:

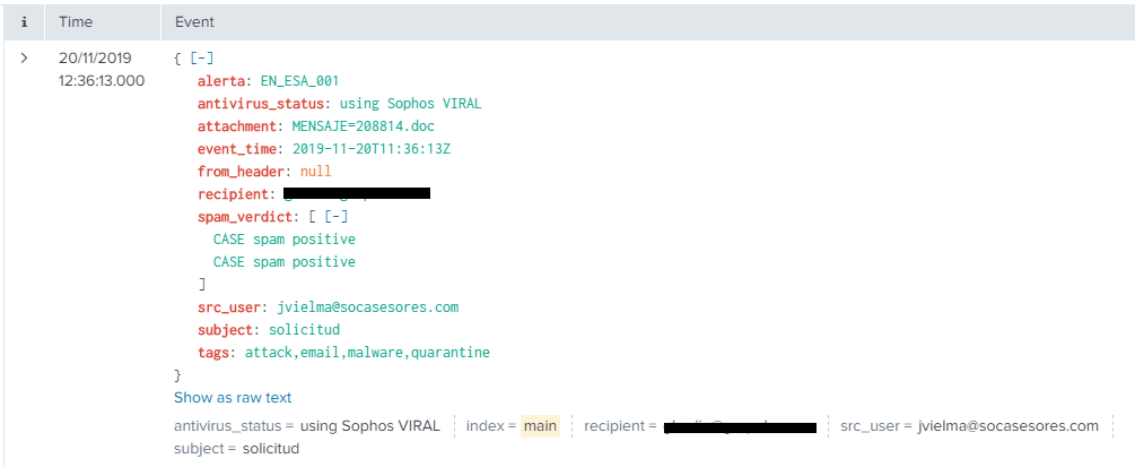

Ilustración 35. Evento EN_ESA_001 indexado en Splunk

```
# Alert EN_ESA_002
es_host: IP_ELASTIC
es_port: 9200
name: EN_ESA_002
description: Seguridad
type: frequency
run_every:
  minutes: 1
buffer_time:
  minutes: 2
realert:
  hours: 0
index: logstash-ironport*
query_key: ["@timestamp","iron_from","ion_to"]
include:
["iron_from","iron_domain_fromheader","iron_to","iron_subject_full","i
ron_attachment","iron_tagging","iron_av","iron_engine","iron_contentfi
lter"]
num_events: 1
```

```
timeframe:
  minutes: 1
filter:
- bool:
        must:
                - query_string:
                                query:      "logger:      finished      AND
iron_contentfilter: URLBADREPUTATION"
alert: post
alert_subject: "EN_ESA_002"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_time: '@timestamp'
   src_user: 'iron_from'
   from_header: 'iron_domain_fromheader'
   recipient: 'iron_to'
   subject: 'iron_subject_full'
   attachment: 'iron_attachment'
   antivirus_status: 'iron_av'
   spam_verdict: 'iron_engine'
   content_filter: 'iron_contentfilter'
   prioridad: '20'
http_post_static_payload:
   alerta: 'EN_ESA_002'
   tags: "attack,email,badreputation,spam,delivery"
http_post_headers:
   X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
   Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
   Accept: "*/*"
```

Una vez activada la alerta, cuando se detecta una ocurrencia se envía a Splunk:

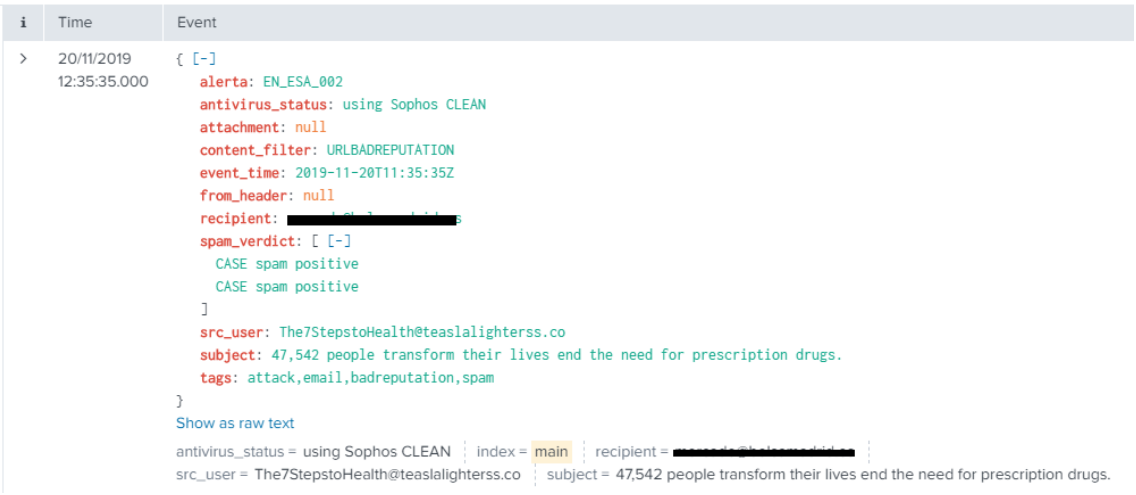

Ilustración 36. Evento EN_ESA_002 indexado dentro de Splunk

Para la normalización de los campos se ha seguido la información contenida dentro de la TA oficial de Cisco ESA que podemos encontrar dentro de Splunkbase.

3.12.5. Alternativa de alertas con Watcher.

En el presente trabajo hemos creado las alertas únicamente usando el correlador ElastAlert. Se considera que para el propósito que queremos realizar, las funcionalidades de Watcher no cumplen nuestras necesidades.

Watcher es una excelente alternativa cuando queremos correlar en una misma alerta varios índices. La mayor problemática que posee este correlador es que el resultado que se obtiene independientemente que sea múltiple, se envía en una única alerta. Es decir, si el resultado de la ejecución son la generación de 4 eventos distintos, en lugar de generar un post por cada uno de estos, lo que hace es generar un único evento con todos los resultados juntos.

Esta forma de darnos el resultado no es válida cuando queremos generar eventos notables de forma individual. Sí que nos será útil en el caso de alertas que devuelvan un único resultado, o el resultado que devuelve aunque sea múltiple, se devuelve un único campo (por ejemplo, un conjunto de hashes).

Veamos a continuación una correlación basada en un Watcher que busca la detección del evento de seguridad de Windows 4625.

```
{
  "trigger": {
    "schedule": {
      "interval": "1m"
    }
  },
  "input": {
    "search": {
      "request": {
        "search_type": "query_then_fetch",
        "indices": [
          "logstash-winlogbeat*"
        ],
        "rest_total_hits_as_int": true,
        "body": {
          "size": 9999,
          "query": {
            "bool": {
              "must": [
                {
                  "range": {
                    "@timestamp": {
                      "gte": "now-1m",
                      "lte": "now"
                    }
                  }
                },
                {
                  "query_string": {
                    "query": "event_id:\"4625\"",
                    "analyze_wildcard": true
                  }
                }
              ]
            }
          }
        }
      }
    }
```

```
  },
  "condition": {
    "compare": {
      "ctx.payload.hits.total": {
        "not_eq": 0
      }
    }
  },
  "actions": {
    "splunk_webhook": {
      "webhook": {
        "scheme": "http",
        "host": "NOMBRE_HOST_SPLUNK",
        "port": 8088,
        "method": "post",
        "path": "/services/collector/raw",
        "params": {},
        "headers": {
          "Authorization":       "Splunk       915512b6-05a9-416d-8ede-
3aa51a320619",
          "X-Splunk-Request-Channel":         "76a98344-8664-4b4e-bc4e-
f0c8dc518688",
          "Content-Type": "application/json"
        },
        "body":                  "{\"user\":\"{{#ctx.payload.hits.hits}}
{{_source.event_data.TargetUserName}}
{{/ctx.payload.hits.hits}}\",\"computer_name\":\"{{#ctx.payload.hits.h
its}} {{_source.computer_name}} {{/ctx.payload.hits.hits}}\"}"
      }
    }
  }
}
```

Como se puede ver en los eventos indexados en Splunk, se envía un único evento con múltiples campos y valores. De esta forma para poder trabajar con el evento se debería parsear para extraer los distintos campos. Esto en sí no es operativo.

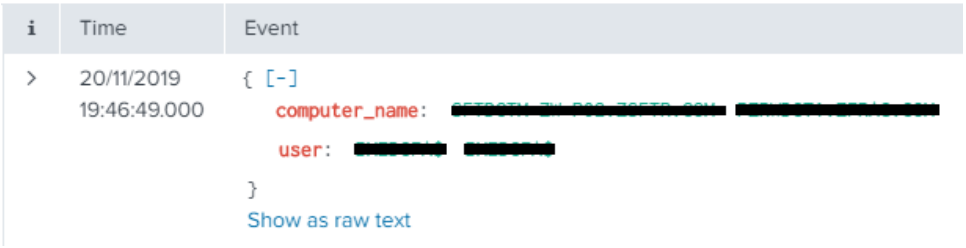

Ilustración 37. Ejemplo de indexación alerta Watcher

Por ello, únicamente vamos a hacer uso de este tipo de correlación en casos muy concretos. Un posible ejemplo que podríamos implementar es si tenemos integrado un feed de inteligencia tipo MISP, un cruce entre IOCs de tipo IP con eventos del FW. De esta forma podríamos detectar IP's atacantes por ejemplo, listadas dentro del MISP y que está teniendo algún tipo de actividad dentro de nuestra red.

Cálculo del espacio en disco (consumo de licencia) de los eventos notables indexados.

Como se ha ido indicando a lo largo del presente trabajo, lo que nos interesa es enviar el mayor número de eventos notables desde el Elastic, hasta Splunk. Y como sabemos bien, el espacio que podemos indexar dentro de Splunk es muy limitado a nivel de licencia. Por tanto, a continuación se va a mostrar la forma de ver qué tamaño se está indexando por día y por sourcetype.

```
index = _internal source = *license_usage.log type = "Usage"
  | eval indexname = if(len(idx) = 0 OR isnull(idx), "(UNKNOWN)", idx)
  | eval sourcetypename = st
  | bin _time span = 1d
  | stats sum(b) as b by _time, pool, indexname, sourcetypename
  | eval GB = round(b/1024/1024/1024, 3)
  | fields _time, indexname, sourcetypename, GB
```

O por índice, o en total:

```
index = _internal source = *license_usage.log type = "Usage"
  | eval indexname = if(len(idx) = 0 OR isnull(idx), "(UNKNOWN)", idx)
  | bin _time span = 1d
  | stats sum(b) as b by _time, pool, indexname
  | eval GB = round(b/1024/1024/1024, 3)
  | fields _time, indexname, GB

index = _internal source = *license_usage.log type = "Usage"
  | bin _time span = 1d
  | stats sum(b) as b by _time, pool
  | eval GB = round(b/1024/1024/1024, 3)
  | fields _time, GB
```

Indicar que el propio sistema controla el consumo de licencia que estamos realizando de forma diaria. En el siguiente enlace podemos ver cómo funciona exactamente esto. El violar repetidamente la licencia puede hacer que tengamos problemas para mantenerla en el caso en que no queramos o podamos ampliarla.

Pero mediante la consulta anterior, podemos adelantarnos ante una posible violación de licencia, alertándonos cuando empecemos a estar cerca de un umbral determinado. De esta forma, podremos ir ajustando el número de eventos notables o su tamaño que se envía.

No se desarrollará más este concepto, pero se deja claro su potencialidad.

3.13. Análisis eventos notables y creación correlación SIEM.

3.13.1. Elementos previos.

Asignación del valor del CKC a los eventos notables.

Acabamos de ver una serie de ejemplos de cómo crear eventos notables. En cada uno de estos eventos, se han definido una serie de etiquetas. Pero aún no están del todo completos. Vamos a identificar cada uno de los eventos notables dentro del CKC y añadírselas a estos. Esto nos ayudará posteriormente para darle la prioridad de tratamiento, y si generarán o no alertas por sí mismas.

Evento Notable	Fase del CKC	Etiqueta
EN_Windows_4624_10	Actions on Objective	AoO
EN_Windows_4625	Exploitation	exploitation
EN_Windows-Wazuh_001	Installation	installation
EN_Windows_explorer	Installation	Installation
EN_CiscoAMP	Exploitation	exploitation
EN_Firewall_001	Exploitation	exploitation
EN_Firewall_002	Reconnaissance ; Command & Control	recon ; C2
EN_ESA_001	Delivery	delivery
EN_ESA_002	Delivery	delivery

Tabla 17. Tabla eventos notables, fase CKC y etiquetas

Después de este cálculo, lo añadimos a las distintos eventos notables dentro de su campo tags.

Cálculo de la prioridad.

Una vez que tenemos definidas todas las etiquetas de los distintos eventos notables, vamos a darle a cada uno de estos un valor numérico de prioridad en la gestión de la alerta. El valor máximo va a ser de 100. La siguiente tabla muestra los rangos:

Rango numérico	Prioridad
80 – 100	Crítica
60 - 80	Alta
40 – 60	Media
20 – 40	Baja
0 - 20	Información

Tabla 18. Prioridades

Las variables van a ser las siguientes:

- Posición dentro del CKC.
- Evento bloqueado o no.
- Número de usuarios o máquinas implicados.
- Evento complejos o de appliances de seguridad.

CKC	Valor
Recon	10
Weapon	10
Delivery	20
Exploitation	30
Installation	30
C2	40
AoO	50

Tabla 19. Valores en función de CKC

Si el evento no es bloqueado, se suma 20 al resultado, sin sobrepasar el máximo permitido de 100.

Si existe la posibilidad de que el evento notable detecte una amenaza para múltiples identidades dentro de la organización, se incrementará su valor en 10 puntos. En todo caso, lo normal es que los eventos notables hagan referencia a eventos unitarios, y la detección de amenazas combinadas o múltiples se produzca ya en las alertas.

Por último se aplica un factor de corrección a los eventos, dependiendo de si se basan en eventos complejos y selectivos-filtrados, si provienen de un appliance de seguridad, o únicamente se trata de un evento más o menos genérico o con alta volumetría en cuanto a su detección. Si son eventos genéricos que necesitan de una posterior correlación o filtrado más en detalle, se le restará 10 puntos a su prioridad de base.

Evento Notable	CKC	Bloqueado	# targets implicados	Gen. / Espec.	prioridad
EN_Windows_4624_10	50	20	0	-10	60
EN_Windows_4625	30	0	0	-10	20
EN_Windows-Wazuh_001	30	20	0	-10	40
EN_Windows_explorer	30	20	0	-10	40
EN_CiscoAMP	30	0/20	0	0	30/50
EN_Firewall_001	30	0/20	0	0	30/50

EN_Firewall_002	10/40	20	0	-10	20/50
EN_ESA_001	20	0	0	0	20
EN_ESA_002	20	0	0	0	20

Tabla 20. Prioridades finales

Como se puede ver, eventos aparentemente genéricos han quedado con un valor de prioridad bastante alto. Esto ha sido principalmente motivado por su posición dentro del CKC. En todo caso, los valores anteriores son solo orientativos. La criticidad de base de cada uno de los eventos notables podrá sufrir alguna corrección específica, antes del envío final.

En concreto, podemos ver que hay algún evento notable con un valor o bien algo alto, o que puede adoptar dos posibles valores finales. Como en esta fase del proyecto de la creación del SIEM no se va a crear una alerta que dispare automáticamente un ticket por dicho valor, no vamos a modificar el valor excesivo de por ejemplo el evento notable EN_Windows_4624_10. En lo anterior se profundizará más adelante. Y para los eventos con dos posibles valores de prioridad, lo que haremos es asignarle su valor mediana, y posteriormente ya dentro de las alertas del SIEM, determinaremos una vez usemos dichos eventos, qué prioridad real tendrá.

Por tanto, los valores de las prioridades finales serán los siguientes:

Evento Notable	prioridad	
EN_Windows_4624_10	60	Alta
EN_Windows_4625	20	Baja
EN_Windows-Wazuh_001	40	Media
EN_Windows_explorer	40	Media
EN_CiscoAMP	40	Media
EN_Firewall_001	40	Media
EN_Firewall_002	35	Baja
EN_ESA_001	20	Baja
EN_ESA_002	20	Baja

Tabla 21. Prioridad numérica y palabra

Por último, una vez que utilicemos los eventos notables para realizar correlaciones más complejas, se incorporará al peso de la prioridad, si estamos detectando amenazas en múltiples máquinas o sobre varios usuarios, y si estos son considerados assets críticos.

Por tanto, ya solo nos queda incluir un nuevo campo dentro de cada uno de los eventos notables creados. Este campo se llamará prioridad, y será un valor numérico calculado en base a la tabla anterior.

Prioridad y SLA.

La prioridad va íntimamente asociada con el tiempo en que se debe gestionar una alerta dada. La prioridad está asociada por tanto, a la criticidad del evento

detectado dentro del marco de la gestión de incidentes de seguridad. Los valores temporales serán los siguientes:

Prioridad	SLA
Crítica	30 min.
Alta	1 hora
Media	4 horas
Baja	12 horas
Información	24 horas

Tabla 22. Prioridad vs SLA

Indicar que, los valores que se acaban de asignar son meramente orientativos. Dependerán fuertemente de los sistemas que se monitoricen y la infraestructura que se esté protegiendo. Son por tanto valores que cada SOC debe de analizar y determinar de forma independiente.

3.13.2. Alertas sistemas Windows.

En esta sección, nos vamos a centrar en poner una serie de ejemplos de alertas que podemos implementar dentro de los eventos notables de Windows. El ejercicio va a estar relacionado con el evento 4625 que analizamos en profundidad dentro de la sección de los eventos notables visto anteriormente.

Vamos a crear siete alertas basadas en el evento 4625. Este evento bien correlado, puede hacer que detectemos acciones sospechosas interesantes.

Nombre de la alerta	Descripción de la alerta
UC01-WIN-001 Multiple Failed Logins from different accounts same host	Alerta que detecta múltiples login fallidos generados por múltiples cuentas dentro de una misma máquina.
UC01-WIN-002 Multiple Failed Logins from same account same host	Alerta que detecta múltiples login fallidos para una misma cuenta, en la misma cuenta.
UC01-WIN-003 Multiple failed logins of non-existing accounts same host	Alerta que detecta múltiples eventos de login fallidos de una cuenta que no existe.
UC01-WIN-004 Multiple failed logins of non-existing account different hosts	Alerta que detecta múltiples eventos fallidos de login de una cuenta no existente en múltiples hosts.
UC01-WIN-005	Alerta que detecta intentos de login fallidos con

Suspicious attempt to logon with Guest account	el usuario Guest (invitado).
UC01-WIN-006 Suspicious Failed Logon error codes	Alerta que detecta intentos fallidos de login con un estatus de error sospechoso.
UC01-WIN-007 Multiple brute-force authentication attempts in different hosts	Alerta creada para detectar intentos de login fallidos de tipo fuerza bruta.

Tabla 23. Alertas Windows evento 4625

Analicemos cada una de estas alertas.

UC01-WIN-001 Multiple Failed Logins from different accounts same host

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC01-WIN-001	Media	Cada 15 minutos	4 horas

Tabla 24. Información de la alerta UC01-WIN-001

La siguiente alerta se encarga de detectar múltiples login fallidos generados por múltiples cuentas dentro de una misma máquina. Por tanto, la correlación dentro de Splunk será la siguiente:

```
index = main sourcetype = _json alerta = EN_Windows_4625 user != *$
user != "-"
| stats dc(user) as distinct_user values(src_host) as src_host
values(user) as user values(Status) as Status values(Sub_Status) as
Sub_Status count values(signature_id) as signature_id values(name) as
Name values(src) as src values(Logon_Type) as Logon_Type
earliest(_time) as first_event latest(_time) as last_event by src_ip
| where distinct_user >= 10
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host
src_ip user Status Sub_Status distinct_user count
```

Como se puede ver en la correlación, lo más destacado es que la agrupación se realiza por la IP de la máquina que genera los eventos 4625, calculándose a su vez el número total de usuarios distintos (distinct_user). Una vez calculado este valor, la alerta se disparará únicamente a partir de la detección de al menos 10 usuarios distintos dentro de la ventana de tiempo establecida dentro de la alerta. Se filtran las cuentas de las máquinas.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	No → 0	Sí/No →	20	55

30/0

Tabla 25. Prioridad de la alerta UC01-WIN-001

En la tabla anterior, hemos podido ver cómo se calcula la prioridad final de una alerta. En primer lugar, tenemos la prioridad base que nos la da el evento notable. Luego se comprueba el número de máquinas posibles afectadas. Este valor depende enteramente de la naturaleza de la alerta, dado que el EN de tipo 4625 es individual. Posteriormente, se analiza si se está monitorizando un elemento crítico en dicha alerta. En este caso al ser una alerta bastante genérica, se pueden llegar a monitorizar tanto máquinas con una prioridad normal, como con prioridad alta. Por tanto, será en el procedimiento donde se determinará la prioridad final (una vez realizado un triaje inicial de la alerta generada por un elemento del equipo de seguridad). Ya por último, se introduce un factor de corrección en función de la propia naturaleza de lo que se esté buscando con la alerta, y del previsible valor de la prioridad final. En este caso en concreto, se ha utilizado dicho factor de corrección.

El resultado final de 55 proviene de la suma de la prioridad de base, del factor de corrección, y de la media surgida de si se está monitorizando un sistema crítico o no.

La prioridad de la alerta se puede hacer “visible” dentro de la propia alerta, en su código. O bien, dependiendo de la herramienta de ticketing que se posea, se puede tener una tabla de equivalencias.

La forma de reflejarlo dentro de la alerta sería de la siguiente forma: UC01-WIN-**3**-001 Todo ello basado en la siguiente tabla:

Prioridad	#
Crítica	1
Alta	2
Media	3
Baja	4
Información	5

Tabla 26. Prioridad dentro del identificador de la alerta UC01-WIN-001

Dado que el presente trabajo no va a ser práctico, en el sentido que no va a estar en producción, se decide crear un CSV de relaciones que posteriormente será usado dentro del Dashboard general, y no indicarlo en la propia nomenclatura de la alerta.

Etiquetado:

También debemos añadir que también se van a crear etiquetas para las alertas. Se van a basar tanto en las etiquetas origen del evento notable, como de otras nuevas surgidas de la propia alerta.

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, failure, windows, security, error, os, exploitation	medium

Tabla 27. Etiquetado de la alerta UC01-WIN-001

Creación y planificación de la alerta dentro de Splunk:

En las siguientes líneas se va a mostrar la forma en que se crea y planifica una alerta dentro de Splunk Enterprise. Se detallará únicamente en la presente alerta a modo de guía.

Como se ha comentado anteriormente, una alerta no es más que una búsqueda planificada. A continuación, vemos la búsqueda dentro de Splunk:

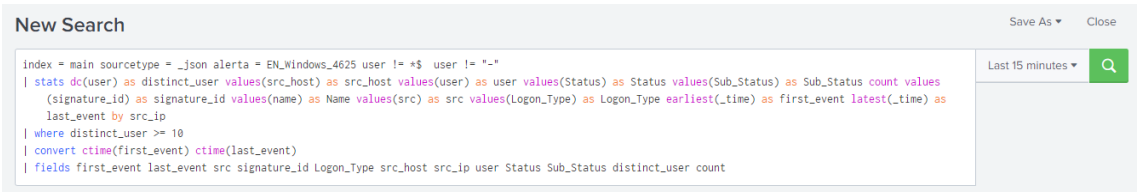

Ilustración 38. Búsqueda dentro de Splunk

Para guardarla como alerta debemos de irnos a Save As > Alert:

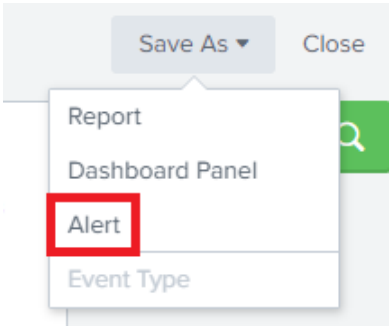

Ilustración 39. Guardado como alerta

Una vez que le hemos indicado que queremos que se guarde como una alerta, nos aparecerá una ventana donde vamos a configurar todos sus parámetros.

Empezaremos por definir tanto su nombre, como el rango de tiempo en que se va a ejecutar, como su planificación a través de un Cron.

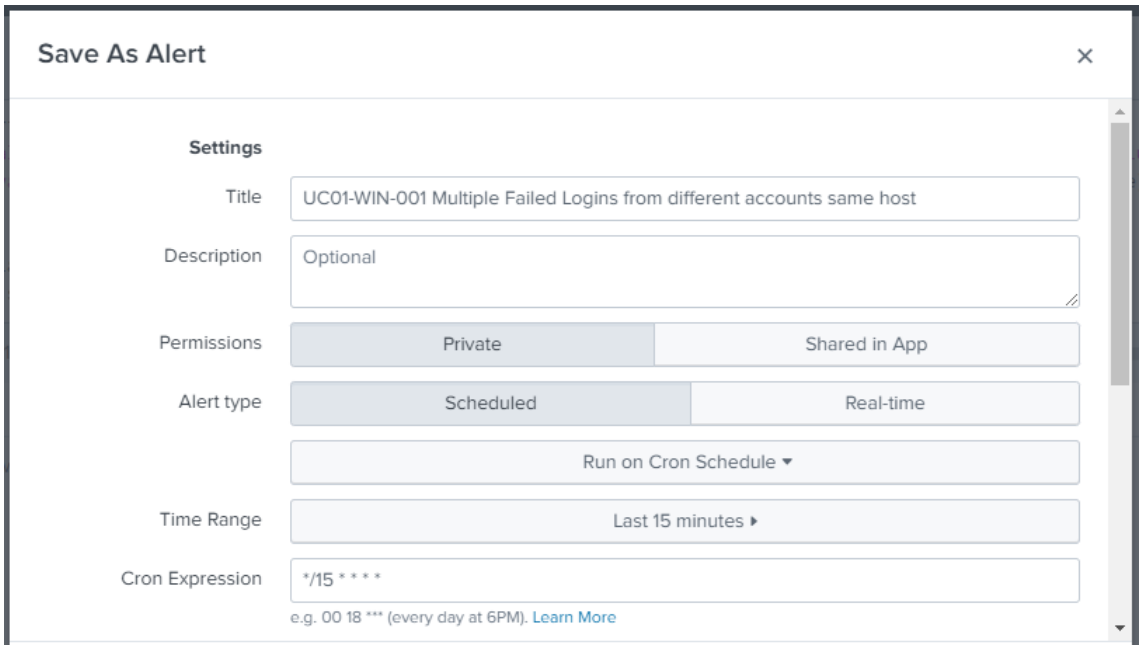

Ilustración 40. Configuración de la alerta (I)

En este caso, se le ha indicado que la ventana de tiempo sea de 15 minutos, al igual que el tiempo de planificado de la alerta. En el presente trabajo no se va a entrar en detalle, pero sí se va a indicar que obviamente estos tiempos son importantes y están relacionados con la prioridad de cada una de las alertas.

Además, dejar indicado, que es normal que si las alertas son pesadas en cuanto a consumo de recursos, la planificación temporal de ejecución sea analizada para que no se solapen entre alertas pesadas y perjudiquen el rendimiento de la máquina sobre la que corre el servicio de Splunk.

A continuación, le vamos a indicar el silenciamiento de la alerta, y si se dispara una alerta por cada uno de los resultados (alerta) o solo una con todos ellos (reporte).

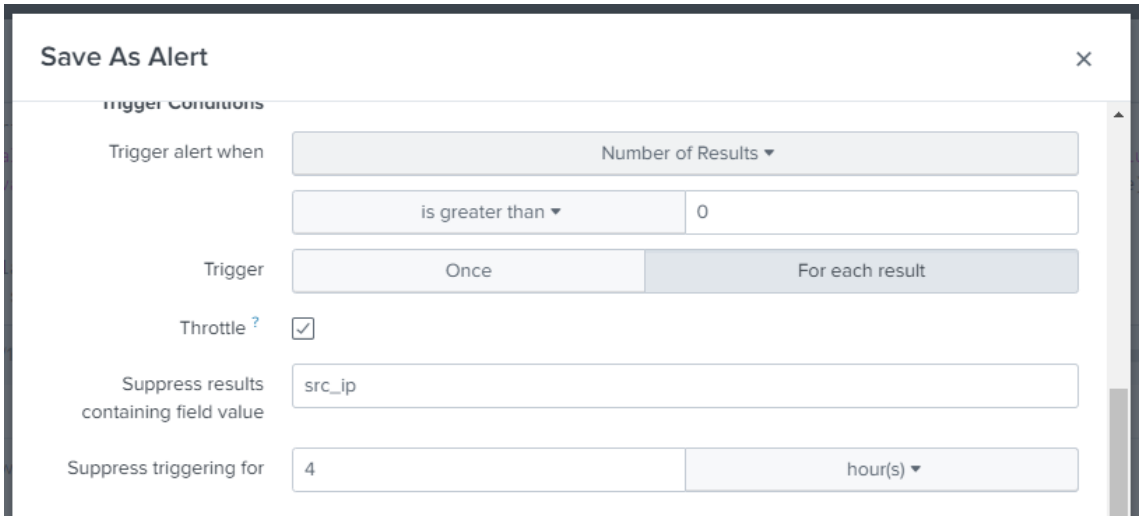

Ilustración 41. Configuración de la alerta (II)

Como se puede ver, se va a ejecutar en modo alerta (For each result), y hay un silenciamiento por el campo de agregación src_ip, durante 4 horas, como se ha indicado anteriormente.

Por último, le indicamos las dos acciones que se van a llevar a cabo una vez que se dispara la alerta.

Por un lado, se enviará un correo a la herramienta de ticketing:

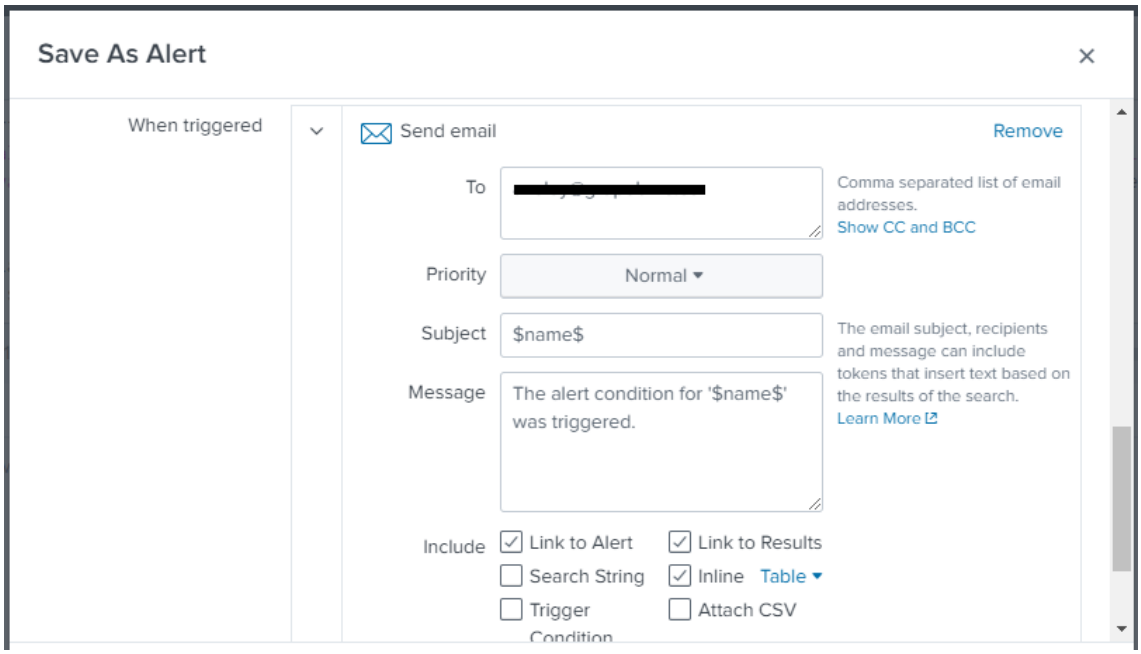

Ilustración 42. Configuración de la alerta (III)

Interesante aquí es el subject que se le asigna al correo. Normalmente en base a la nomenclatura de la alerta, dentro de la herramienta de ticketing se puede procesar de una forma u otra la alerta.

Dentro del correo aparecerán también, el link a la alerta que acaba de generarse, otro link hacia los resultados y la propia tabla con el resultado.

La otra acción consiste en la generación de un evento resultado de la alerta que acaba de dispararse.

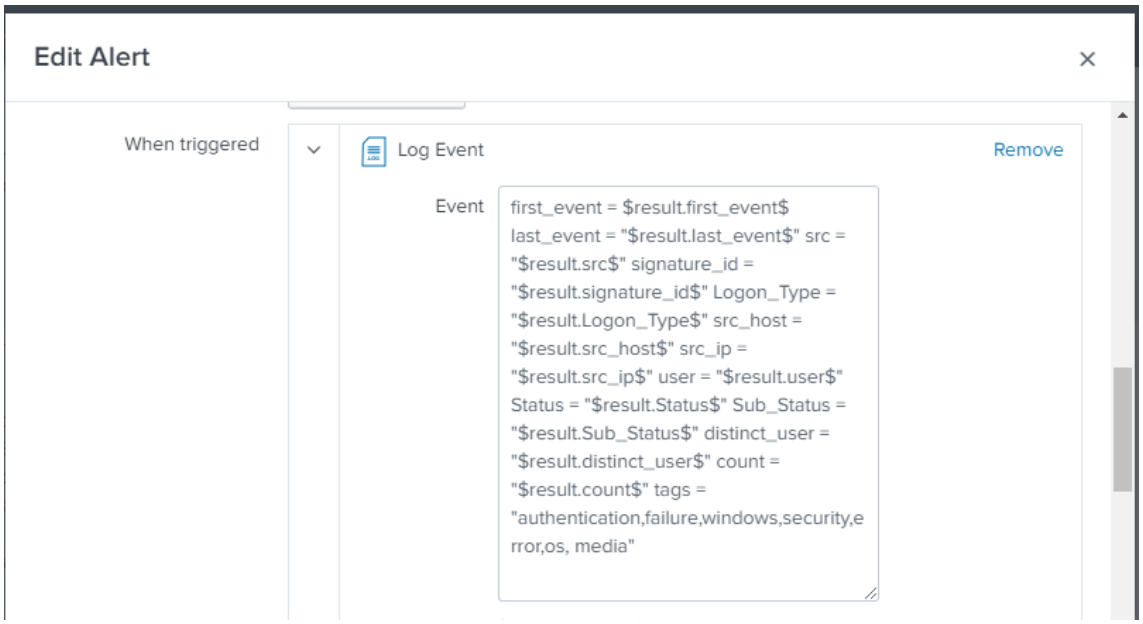

Ilustración 43. Configuración del Log Event (I)

En el campo Event, se definen todos los valores del log que queremos indexar.

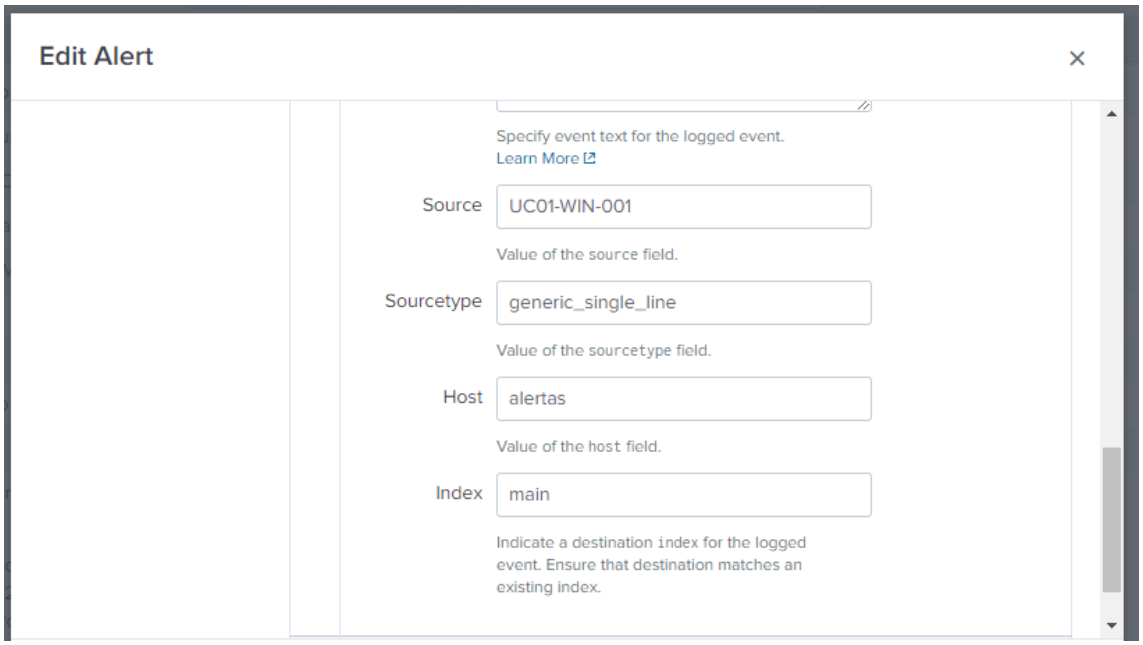

Ilustración 44. Configuración del Log Event (II)

Se define para finalizar, los metadatos del evento.

La idea que se está pretendiendo con este evento es poder tanto hacer estadísticas (KPIs), como incluso, utilizar los datos para generar otras alertas o reportes.

A continuación, se muestra como se ve la alerta dentro de la sección de alertas de la APP donde la hemos creado.

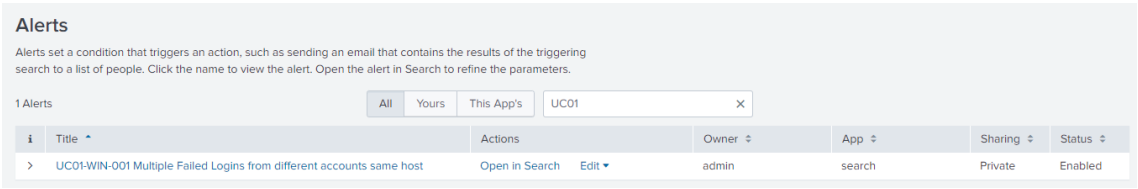

Ilustración 45. Vemos la alerta resumida

UC01-WIN-002 Multiple Failed Logins from same account same host

Identificador alerta	prioridad	Planificación temporal	Silenciamiento
UC01-WIN-002	Media	Cada 15 minutos	4 horas

Tabla 28. Información de la alerta UC01-WIN-002

La presente alerta, se encarga de detectar múltiples intentos fallidos de un mismo usuario, independientemente de la máquina origen donde se produzcan estos.

El código de la correlación es el siguiente:

```
index = main sourcetype = _json alerta = EN_Windows_4625 user != *$
| stats count dc(src_ip) as distinct_src_ip values(src_host) as
src_host values(src_ip) as src_ip values(Status) as Status
values(Sub_Status) as Sub_Status values(signature_id) as signature_id
values(name) as Name values(src) as src values(Logon_Type) as
Logon_Type earliest(_time) as first_event latest(_time) as last_event
by user
| where count >= 20 AND distinct_src_ip = 1
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host
src_ip user Status Sub_Status count
```

A nivel de correlación, destacar que la agrupación se realiza por el usuario, y que el umbral marcado es de al menos 20 intentos fallidos en la ventana de tiempo indicado en la alerta. Como filtro se ha descartado las cuentas de máquina.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	0	Sí/No → 30/0	20	55

Tabla 29. Prioridad de la alerta UC01-WIN-002

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, failure, windows, security, error, os, exploitation	medium

Tabla 30. Etiquetado de la alerta UC01-WIN-002

UC01-WIN-003 Multiple failed logins of non-existing accounts same host

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC01-WIN-003	Media	Cada 15 minutos	4 horas

Tabla 31. Información de la alerta UC01-WIN-003

Con esta alerta se intenta detectar múltiples intentos fallidos de cuentas que no existen, dentro de un mismo host.

La correlación de la alerta es la siguiente:

```
index = main sourcetype = _json alerta = EN_Windows_4625 Sub_Status =
0xC0000064 user != *$ src_ip != ::1 src_ip != -
| stats count dc(user) as distinct_user values(src_host) as src_host
values(user) as user values(Status) as Status values(Sub_Status) as
Sub_Status values(signature_id) as signature_id values(name) as Name
values(src) as src values(Logon_Type) as Logon_Type earliest(_time) as
first_event latest(_time) as last_event by src_ip
| where distinct_user >= 6
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host
src_ip user Status Sub_Status distinct_user count
```

Como se puede ver en el código, se agrega por la IP de origen donde se generan los eventos. Se cuentan el número total de usuarios distintos detectados, siendo al menos de 6 para que la alerta se dispare. Dentro de los filtros de la alerta, destacar que se busca el valor del campo Sub_Status igual a 0xC0000064 (user name does not exist).

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	0	Sí/No → 30/0	20	55

Tabla 32. Prioridad de la alerta UC01-WIN-003

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, failure, windows, security, error, os, exploitation	medium

Tabla 33. Etiquetado de la alerta UC01-WIN-003

UC01-WIN-004 Multiple failed logins of non-existing account different hosts

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC01-WIN-004	Alta	Cada 5 minutos	1 hora

Tabla 34. Información de la alerta UC01-WIN-004

La alerta 004 detecta múltiples intentos de login fallidos de una cuenta no existente en diferentes máquinas.

Correlación:

```
index = main sourcetype = _json alerta = EN_Windows_4625 Sub_Status = 0xC0000064 user != *$ user != "-" src_ip != ::1
| stats dc(src_ip) as distinct_src_ip values(src_host) as src_host values(src_ip) as src_ip values(Status) as Status values(Sub_Status) as Sub_Status values(signature_id) as signature_id values(name) as Name values(src) as src values(Logon_Type) as Logon_Type earliest(_time) as first_event latest(_time) as last_event count by user
| where distinct_src_ip >= 3
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host src_ip user Status Sub_Status distinct_src_ip count
```

La agrupación se realiza por nombre de usuario que no existe, y se calcula el número total de máquinas distintas donde aparece este evento. Este número debe de ser al menos de 3 en la ventana de tiempo especificada. Dentro de los filtros, destacar el mismo valor que el de la alerta 003 de Windows.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	10	Sí/No → 30/0	20	65

Tabla 35. Prioridad de la alerta UC01-WIN-004

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, failure, windows, security, error, os, exploitation	high

Tabla 36. Etiquetado de la alerta UC01-WIN-004

UC01-WIN-005 Suspicious attempt to logon with Guest account

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC01-WIN-005	Alta	Cada 5 minutos	1 hora

Tabla 37. Información de la alerta UC01-WIN-005

La alerta 005 busca detectar múltiples intentos fallidos de la cuenta invitado.

Correlación:

```
index = main sourcetype = _json alerta = EN_Windows_4625 user = "guest"
| stats values(src_host) as src_host values(src_ip) as src_ip values(Status) as Status values(Sub_Status) as Sub_Status count values(signature_id) as signature_id values(src) as src values(Logon_Type) as Logon_Type earliest(_time) as first_event latest(_time) as last_event by user
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host src_ip user Status Sub_Status count
```

La correlación es más simple que las anteriores. Indicar que se pueden detectar dentro de la alerta, eventos de un único host, o de varios. Esto último hará que su tratamiento pueda ser distinto a nivel de severidad y tiempo de respuesta.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	10	Sí/No → 30/0	20	65

Tabla 38. Prioridad de la alerta UC01-WIN-005

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, failure, windows, security, error, os, exploitation	high, guest

Tabla 39. Etiquetado de la alerta UC01-WIN-005

Esta alerta nos va a permitir crear otras dos distintas. Nos vamos a basar en que la cuenta guest dentro de los sistemas Windows están por defecto desactivadas. Por tanto, se pueden crear las dos siguientes alertas:

Nombre de la alerta	Descripción de la alerta
UC01-WIN-008 logon attempt with Guest account – Error Code 0xC000006A	Alerta que detecta errores de acceso de la cuenta Guest dentro de una máquina, cuyo código de error nos indique que la cuenta está habilitada.
UC01-WIN-009 multiple failed login attempts successful login with Guest account	Alerta que detecta intentos fallidos de acceso con la cuenta Guest, junto a un evento de login satisfactorio para dicha cuenta y máquina.

Tabla 40. Nuevas alertas UC01-WIN-008 y UC01-WIN-009

Más adelante, se verán ambas alertas.

UC01-WIN-006 Suspicious Failed Logon error codes

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC01-WIN-006	Media	Cada 15 minutos	4 horas

Tabla 41. Información de la alerta UC01-WIN-006

Esta alerta busca detectar códigos de error especiales:

Código de error	Descripción
0xC0000072	Intento de inicio de sesión de usuario en cuenta deshabilitado por el administrador.
0xC000006F	Intento de inicio de sesión de usuario fuera de horas autorizadas.
0xC0000070	Intento de inicio de sesión de usuario desde una estación de trabajo no autorizada
0xC0000413	Error de inicio de sesión: el equipo en el que está iniciando sesión está protegido por un firewall de autenticación. La cuenta especificada no puede autenticarse en el equipo.
0xC000018C	Error en la solicitud de inicio de sesión porque no se pudo realizar la relación de confianza entre el dominio principal y el dominio de confianza.

Tabla 42. Códigos de error

Correlación:

```
index = main sourcetype = _json alerta = EN_Windows_4625 (Sub_Status =
"*C0000072*" OR Sub_Status = "*C000006F*" OR Sub_Status = "*C0000070*"
OR Sub_Status = "*C0000413*" OR Sub_Status = "*C000018C*")
| stats values(src_host) as src_host values(src_ip) as src_ip
values(Status) as Status values(Sub_Status) as Sub_Status count
values(signature_id) as signature_id values(src) as src
values(Logon_Type) as Logon_Type earliest(_time) as first_event
latest(_time) as last_event by user
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host
src_ip user Status Sub_Status count
```

La correlación se realiza en base a un usuario particular. En la alerta podremos agrupar para el mismo usuario múltiples Sub_Status, y que se produzcan dichos eventos en una o más máquinas. Todo ello influirá a la hora de tratar el incidente.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	10	Sí/No → 30/0	10	55

Tabla 43. Prioridad de la alerta UC01-WIN-006

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, failure, windows, security, error, os, exploitation	medium

Tabla 44. Etiquetado de la alerta UC01-WIN-006

UC01-WIN-007 Multiple brute-force authentication attempts in different hosts

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC01-WIN-007	Alta	Cada 5 minutos	1 hora

Tabla 45. Información de la alerta UC01-WIN-007

La alerta 007 busca detectar un ataque de fuerza bruta basándose en el nombre del mismo usuario en múltiples hosts.

Correlación:

```
index = main sourcetype = _json alerta = EN_Windows_4625 user != *$
src_ip != ::1
| stats dc(src_ip) as distinct_src_ip values(src_host) as src_host
values(src_ip) as src_ip values(Status) as Status values(Sub_Status)
```

```
as Sub_Status count values(signature_id) as signature_id values(src)
as src values(Logon_Type) as Logon_Type earliest(_time) as first_event
latest(_time) as last_event by user
| where distinct_src_ip >= 20
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host
src_ip user Status Sub_Status distinct_src_ip count
```

Como se puede ver en la correlación, se agrupa por el nombre de usuario, contando el número de hosts distintos. Para que salte la alerta, debe de haber al menos 20 hosts distintos en la ventana de tiempo indicada por la alerta.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	10	Sí/No → 30/0	20	65

Tabla 46. Prioridad de la alerta UC01-WIN-007

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, failure, windows, security, error, os, exploitation	high, bruteforce

Tabla 47. Etiquetado de la alerta UC01-WIN-007

UC01-WIN-008 Logon attempt with Guest account – Error Code 0xC000006A

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC01-WIN-008	Alta	Cada 5 minutos	1 hora

Tabla 48. Información de la alerta UC01-WIN-008

La presente alerta proviene del análisis de la 005. En éste, vimos que dado que la cuenta Guest aparece deshabilitada por defecto, es de gran importancia el detectar un intento de acceso que genere como error que dicha cuenta está habilitada. Todo ello producirá que se tenga que crear una nueva alerta, y que a su vez, se deba introducir un filtro dentro de la alerta 005.

Veamos la pequeña modificación que deberíamos realizar en la alerta UC01-WIN-005:

```
index = main sourcetype = _json alerta = EN_Windows_4625 user =
"guest" Sub_Status ¡= "*C000006A*"
| stats values(src_host) as src_host values(src_ip) as src_ip …
```

Como se ve, únicamente hace falta incluir un pequeño filtro que excluya el código de error que vamos a buscar en la presente alerta.

Ahora sí, veamos el código de la nueva alerta.

```
index = main sourcetype = _json alerta = EN_Windows_4625 user = "guest" Sub_Status = "*C000006A*"
| stats values(src_host) as src_host values(Status) as Status values(Sub_Status) as Sub_Status count values(signature_id) as signature_id values(src) as src values(Logon_Type) as Logon_Type earliest(_time) as first_event latest(_time) as last_event by user, src_ip
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host src_ip user Status Sub_Status count
```

Simplemente es modificar un poco la alerta 005, filtrando por el valor que deseamos, y agrupando por usuario y máquina de origen. Con que se produzca un único evento, se disparará la alerta. Indicar que no se correla a través de una tabla (table), dado que podría generarse múltiples eventos en un período corto de tiempo y saturar la herramienta de ticketing si no se implementase un correcto silenciamiento. Además, aunque se tenga un silenciamiento de alerta, esta forma es mejor para que la alerta sea más aproximada a la realidad.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	0	Sí/No → 30/0	30	65

Tabla 49. Prioridad de la alerta UC01-WIN-008

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, failure, windows, security, error, os, exploitation	high, guest

Tabla 50. Etiquetado de la alerta UC01-WIN-008

Otra alerta que no se va a implementar en detalle, pero que sí debería de aparecer dentro de la librería de casos de uso, es la habilitación de la cuenta guest en una máquina. La lógica sería la siguiente: La idea es dado que se va a monitorizar el evento 4722 (A user account was enabled), se podría crear una alerta cuyo filtro fuese la acción de habilitación de la cuenta invitado en una máquina.

```
index = main sourcetype = _json alerta = EN_Windows_4722 user = "guest"
…
```

De esta forma, podríamos detectar una posible violación de la política de seguridad corporativa, y a partir de aquí, monitorizar toda la actividad relacionada.

UC01-WIN-009 Multiple failed login attempts successful login with Guest account

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC01-WIN-009	Crítica	Cada minuto	30 minutos

Tabla 51. Información de la alerta UC01-WIN-009

Como se comentó anteriormente, esta alerta proviene de las posibilidades que ha abierto la alerta 005. Esta alerta nos va a llevar tener que implementar un nuevo evento notable del lado del Elastic. Y es que, vamos a tener que monitorizar el evento de Windows 4624 para el usuario Guest. Una vez que lo estemos indexando dentro de Splunk, haremos la correlación junto al evento notable 4625.

Evento notable dentro del ElastAlert:

Nombre de la alerta	Descripción de la alerta
EN_Windows_4624_Guest	Alerta que se dispara cuando se detecta la generación de un evento de tipo 4624 dentro de un sistema Windows, y cuya cuenta es la del usuario Guest

Tabla 52. Evento notable 4624 Guest

A continuación podemos ver el código de la alerta creada:

```
#Alert EN_Windows_4624_Guest
_host: IP_ELASTIC
run_every:
  minutes: 1
bufferime:
  minutes: 2
es_port: 9200
_source_enabled: true
name: EN_Windows_4624_Guest
description: Seguridad
type: frequency
query_key: "@timestamp"
index: logstash-winlogbeat-*
num_events: 1
include:
["event_data.IpAddress","event_data.LogonType","event_data.TargetUserN
ame","event_data.WorkstationName","event_id","log_name","computer_name
"]
realert:
  hours: 0
timeframe:
  minutes: 1
filter:
        - query:
```

```
            query_string:
                              query:        "event_id:4624        AND
'event_data.TargetUserName: \"guest\""
alert: post
alert_subject: " EN_Windows_4624_Guest"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_type: '@timestamp'
   src_ip: 'event_data.IpAddress'
   Logon_Type: 'event_data.LogonType'
   user: 'event_data.TargetUserName'
   src_host: 'event_data.WorkstationName'
   signature_id: 'event_id'
   Log_Name: 'log_name'
   src: 'computer_name'
http_post_static_payload:
   alerta: 'EN_Windows_4624_Guest'
   tags: "authentication,windows,security,os,suspicious"
http_post_headers:
X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
Accept: "*/*"
```

Una vez que ya tenemos ambos eventos indexándose dentro de Splunk, pasamos a realizar la correlación dentro de Splunk.

```
index = main sourcetype = _json (alerta = EN_Windows_4625 OR alerta =
EN_Windows_4624_Guest) user = "guest"
| stats dc(alerta) as dc_alertas values(src_host) as src_host
values(Status) as Status values(Sub_Status) as Sub_Status count
values(signature_id) as signature_id values(src) as src
values(Logon_Type) as Logon_Type earliest(_time) as first_event
latest(_time) as last_event by user, src_ip
| where dc_alertas = 2
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host
src_ip user Status Sub_Status count
```

La correlación es bastante sencilla. Se busca que para el usuario guest existan ambos eventos dentro de la ventana de tiempo de búsqueda.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	0	30	30	80

Tabla 53. Prioridad de la alerta UC01-WIN-009

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, windows, security, os,	critical, guest,

exploitation	success, access

Tabla 54. Etiquetado de la alerta UC01-WIN-009

Incluso se podría crear una alerta para el caso en que saltase este evento con un único evento 4624. Podría ser interesante el detectar una mala praxis a nivel de administración. La correlación sería muy sencilla:

```
index = main sourcetype = _json alerta = EN_Windows_4624_Guest
| stats count values(src_host) as src_host values(signature_id) as signature_id values(src) as src values(Logon_Type) as Logon_Type earliest(_time) as first_event latest(_time) as last_event by user, src_ip
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host src_ip user count
```

UC01-WIN-010 Multiple failed login attempts successful login with admin account

Nombre de la alerta	Descripción de la alerta
UC01-WIN-010 multiple failed login attempts successful login with admin account	Alerta que lo que detecta son múltiples intentos de acceso con una cuenta de administrador, con un posterior acceso satisfactorio.

Tabla 55. Descripción de la alerta UC01-WIN-010

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC01-WIN-010	Crítica	Cada minuto	30 minutos

Tabla 56. Información de la alerta UC01-WIN-010

Al implementar la alerta relacionada con el usuario guest, hemos podido ver la potencialidad que posee el juntar en una misma búsqueda, los dos eventos anteriores. El problema del sistema que estamos implementando es que la volumetría como se vio del evento 4624 es excesiva. Se puede ir filtrando hasta reducir considerablemente dicho volumen, pero aun así son demasiados eventos. Existen dos posibles aproximaciones, la primera es basarse únicamente en los eventos 4625 dentro de Splunk, para posteriormente consultar dentro del Elastic si existe algún 4624 posterior. Siendo lo anterior bastante manual y poco efectivo (piénsese en que podría tardarse demasiado tiempo en detectar este comportamiento).

La otra aproximación sería el filtrar mucho más los eventos 4624 que se envían a Splunk. Una buena aproximación sería el enviar únicamente los login de una serie de cuentas privilegiadas como podrían ser las cuentas de los administradores de dominio. Bajo esta premisa, podríamos correlar intentos de

acceso fallidos con un acceso posterior de dichas cuentas, y darle a esta alerta una prioridad crítica.

La implementación de dicha alerta sería la siguiente. Se crearía en primer lugar los eventos notables dentro de Elastic relacionados con el evento 4624. Estos eventos se enviarían a Splunk, y ahí haríamos la correlación junto a los eventos 4625 que ya se están enviando.

Evento notable dentro del ElastAlert:

Nombre de la alerta	Descripción de la alerta
EN_Windows_4624_admin	Evento notable que se envía a Splunk cuando se detecta un evento 4624 de una cuenta de administrador.

Tabla 57. Evento notable 4624 Admin

A continuación podemos ver el código de la alerta creada:

```
#Alert EN_Windows_4624_Admin
_host: IP_ELASTIC
run_every:
  minutes: 1
bufferime:
  minutes: 2
es_port: 9200
_source_enabled: true
name: EN_Windows_4624_Admin
description: Seguridad
type: frequency
query_key: "@timestamp"
index: logstash-winlogbeat-*
num_events: 1
include:
["event_data.IpAddress","event_data.LogonType","event_data.TargetUserN
ame","event_data.WorkstationName","event_id","log_name","computer_name
"]
realert:
  hours: 0
timeframe:
  minutes: 1
filter:
        - query:
                query_string:
                                query:        "event_id:4624        AND
event_data.TargetUserName: *administrator*"
alert: post
alert_subject: " EN_Windows_4624_Admin"
http_post_url:
"http://NOMBRE_SPLUNK_SEARCHEAD::8088/services/collector/raw"
http_post_all_values: False
http_post_payload:
   event_type: '@timestamp'
   src_ip: 'event_data.IpAddress'
   Logon_Type: 'event_data.LogonType'
   user: 'event_data.TargetUserName'
   src_host: 'event_data.WorkstationName'
   signature_id: 'event_id'
```

```
   Log_Name: 'log_name'
   src: 'computer_name'
http_post_static_payload:
   alerta: 'EN_Windows_4624_Admin'
   tags: "authentication,windows,security,os,suspicious"
http_post_headers:
X-Splunk-Request-Channel: "76a98344-8664-4b4e-bc4e-f0c8dc518688"
Authorization: "Splunk c619300b-951a-45d1-8bd2-4fe971aeab4e"
Accept: "*/*"
```

Indicar en todo caso que habría que analizar el volumen de eventos que se generan a lo largo del tiempo de este tipo, y cuánto espacio supondría para la licencia contratada en Splunk.

Por último, veamos la correlación dentro de Splunk para esta alerta.

```
index = main sourcetype = _json (alerta = EN_Windows_4624_Admin OR
alerta = EN_Windows_4625) user = *administrator*
| stats dc(alerta) as dc_alertas count values(src_host) as src_host
values(signature_id) as signature_id values(src) as src
values(Logon_Type) as Logon_Type earliest(_time) as first_event
latest(_time) as last_event by user, src_ip
| where dc_alertas = 2
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host
src_ip user count
```

Como se puede ver, la idea es crear una correlación agrupando por usuario y origen, y detectando que existe tanto eventos de tipo 4625, como de 4624. Se filtra originalmente por el campo usuario deseado.

En caso de no tener una nomenclatura estándar para los usuarios administradores, se puede meter todos ellos dentro de una lista en un CSV e importarlo dentro de la consulta:

```
index = main sourcetype = _json (alerta = EN_Windows_4624_Admin OR
alerta = EN_Windows_4625)
| search [| inputlookup administrators.csv]
| stats dc(alerta) as dc_alertas count values(src_host) as src_host
values(signature_id) as signature_id values(src) as src
values(Logon_Type) as Logon_Type earliest(_time) as first_event
latest(_time) as last_event by user, src_ip
| where dc_alertas = 2
| convert ctime(first_event) ctime(last_event)
| fields first_event last_event src signature_id Logon_Type src_host
src_ip user count
```

En el código anterior, podemos ver que la comprobación de los usuarios se realiza a través de la importación de un CSV.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	0	30	30	80

Tabla 58. Prioridad de la alerta UC01-WIN-010

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, windows, security, os, exploitation	critical, admin, success, access

Tabla 59. Etiquetado de la alerta UC01-WIN-010

Análisis de las alertas desplegadas.

Después de ver todas las alertas, hemos llegado a sacar hasta un total de 12 posibles alertas a implementar en un primer análisis para el evento 4625.

3.13.3. Alertas Antivirus.

Habíamos visto que a nivel de eventos del antivirus se enviaban todos los eventos que permitía la API del EDR. Para desarrollar las alertas vamos a tener que echar mano de la típica lógica de detección por volumetría, o detecciones en distintos hosts.

La siguiente tabla enumera las posibles alertas que podemos implementar:

Nombre de la alerta	Descripción de la alerta
UC02-AV-001 Several detection same host	Alerta que detecta dentro de un mismo host, múltiples detecciones de malware.
UC02-AV-002 Several detection same malware multiple hosts	Alerta que detecta múltiples detecciones del mismo malware, en múltiples hosts.
UC02-AV-003 Several detections same malware same host	La alerta detecta múltiples detecciones del mismo malware dentro del mismo hosts.
UC02-AV-004 Virus detected and not removed	Alerta que detecta la no eliminación de un malware detectado dentro de un hosts.
UC02-AV-005 AV Report	Alerta en formato reporte, que se envía todos los días para su revisión.

Tabla 60. Alertas del EDR

Dentro de las alertas anteriores, el parámetro fundamental va a ser la correcta definición del umbral por el cuál una alerta va a saltar. Personalmente suelo ser en este aspecto bastante conservador en el sentido de que mi umbral para estas alertas es bajo. En todo caso, la experiencia del día a día irá demarcando dicho umbral.

UC02-AV-001 Several detection same host

ID alerta	prioridad	Planificación temporal	silenciamiento
UC02-AV-001	Alta	Cada 5 minutos	1 hora

Tabla 61. Información de la alerta UC02-AV-001

Esta alerta lo que busca es detectar dentro una misma máquina, múltiples eventos de malware en un período de tiempo dado. Se va a considerar como umbral de disparo de la alerta, que se detecte un al menos 3 eventos distintos. Eventos distintos, dado que para el mismo malware se va a implementar la alerta 003.

A continuación, podemos ver la alerta que se implementa en Splunk:

```
index = main sourcetype = _json alerta = EN_CiscoAMP (category = "Threat Detected" OR category = "Executed malware" OR category = "Malicious Activity Detection" OR category = "Threat Quarantined")
| stats values(tags) as tags values(signature) as signature values(category) as category values(action) as action values(file_name) as file_name values(file_hash) as file_hash earliest(_time) as first_event latest(_time) as last_event count by dest
| where count > 2
| sort - count
| convert ctime(first_event) ctime(last_event)
| makemv delim="," tags
| fields first_event last_event dest signature category action file_name file_hash count tags
```

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
40	0	0	20	60

Tabla 62. Prioridad de la alerta UC02-AV-001

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
malware, amp, attack, alert, exploitation	high

Tabla 63. Etiquetado de la alerta UC02-AV-001

UC02-AV-002 Several detection same malware multiple hosts

ID alerta	prioridad	Planificación temporal	silenciamiento
UC02-AV-002	Crítica	Cada minuto	n/a

Tabla 64. Información de la alerta UC02-AV-002

La alerta detecta un mismo malware en múltiples host en una ventana de tiempo dada. Como este hecho suele ser raro, el umbral que se ha decidido usar para la alerta ha sido que a partir de dos hosts distintos, se dispare la alerta.

Alerta en Splunk:

```
index = main sourcetype = _json alerta = EN_CiscoAMP (category = "Threat Detected" OR category = "Executed malware" OR category = "Malicious Activity Detection" OR category = "Threat Quarantined")
| stats values(tags) as tags values(signature) as signature values(category) as category values(action) as action values(file_name) as file_name earliest(_time) as first_event latest(_time) as last_event values(dest) as dest dc(dest) as cardinality_dest count by file_hash
| where cardinality_dest > 1
| sort - count
| convert ctime(first_event) ctime(last_event)
| makemv delim="," tags
| fields first_event last_event dest signature category action file_name file_hash count tags
```

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
40	20	0	20	80

Tabla 65. Prioridad de la alerta UC02-AV-002

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
malware, amp, attack, alert,	critical, propagation

exploitation

Tabla 66. Etiquetado de la alerta UC02-AV-002

UC02-AV-003 Several detections same malware same host

ID alerta	prioridad	Planificación temporal	silenciamiento
UC02-AV-003	Alta	Cada 5 minutos	1 hora

Tabla 67. Información de la alerta UC02-AV-003

La alerta salta cuando se ha detectado dentro de una máquina dada, el mismo malware múltiples veces. Para este caso, la ventana de tiempo se debe de ampliar, dado que se está buscando persistencia. Se decide a su vez que el umbral del disparo de la alerta sea de al menos dos.

Alerta dentro de Splunk:

```
index = main sourcetype = _json alerta = EN_CiscoAMP (category = "Threat Detected" OR category = "Executed malware" OR category = "Malicious Activity Detection" OR category = "Threat Quarantined")
| stats values(tags) as tags values(signature) as signature values(category) as category values(action) as action values(file_name) as file_name earliest(_time) as first_event latest(_time) as last_event count by file_hash, dest
| where count > 1
| sort - count
| convert ctime(first_event) ctime(last_event)
| makemv delim="," tags
| fields first_event last_event dest signature category action file_name file_hash count tags
```

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
40	0	0	20	60

Tabla 68. Prioridad de la alerta UC02-AV-003

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
malware, amp, attack, alert, exploitation	high, persistence

Tabla 69. Etiquetado de la alerta UC02-AV-003

UC02-AV-004 Virus detected and not removed

ID alerta	prioridad	Planificación temporal	silenciamiento
UC02-AV-004	Crítica	Cada minuto	n/a

Tabla 70. Información de la alerta UC02-AV-004

La alerta se encarga de detectar la detección de un malware por parte del EDR y que no ha sido posible el eliminarlo o ponerlo en cuarentena.

Correlación en Splunk:

```
index = main sourcetype = _json alerta = EN_CiscoAMP category =
"Quarantine Failure"
| stats values(tags) as tags values(signature) as signature
values(category) as category values(action) as action
values(file_name) as file_name earliest(_time) as first_event
latest(_time) as last_event count values(file_hash) as file_hash by
dest
| sort - count
| convert ctime(first_event) ctime(last_event)
| makemv delim="," tags
| fields first_event last_event dest signature category action
file_hash count tags
```

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
40	0	30	20	90

Tabla 71. Prioridad de la alerta UC02-AV-004

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
malware, amp, attack, alert, exploitation	critical, not-removed

Tabla 72. Etiquetado de la alerta UC02-AV-004

UC02-AV-005 AV Report

ID alerta	prioridad	Planificación temporal	silenciamiento
UC02-AV-005	Alta	Cada 5 minutos	1 hora

Tabla 73. Información de la alerta UC02-AV-005

Esta alerta que es en realidad un reporte envía todos los eventos del día anterior a la herramienta de ticketing para que sea analizado, dado que en éste aparecen todas los eventos relevantes, y por tanto, se pueden detectar cosas interesantes, o que no entrasen dentro del radar de las alertas anteriores.

Correlación dentro de Splunk:

```
index = main sourcetype = _json alerta = EN_CiscoAMP
| makemv delim="," tags
| table _time dest dest_ip category signature action file_name
file_hash tags
```

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
40	10	0	10	60

Tabla 74. Prioridad de la alerta UC02-AV-005

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
malware, amp, attack, alert, exploitation	high, report

Tabla 75. Etiquetado de la alerta UC02-AV-005

Se comentó en su momento, que poseíamos algo de licencia extra dentro de Splunk, y que iba a ser usado para indexar alguna que otra fuente. Dentro de la infraestructura desplegada existe un segundo antivirus (un EPP clásico). En base a esto, vamos a desarrollar una sexta alerta en la que se van a detectar en una misma máquina, múltiples amenazas (al menos dos) provenientes de ambos sistemas antivirus.

UC02-AV-006 Threat detect multi AV

ID alerta	prioridad	Planificación temporal	silenciamiento
UC02-AV-006	Crítica	Cada hora	n/a

Tabla 76. Información de la alerta UC02-AV-006

Correlación dentro de Splunk:

```
((index = antivirus sourcetype = "cisco:amp:event" action = "Threat
Detected" OR (index = main sourcetype = _json alerta = EN_CiscoAMP))
| stats values(signature) as signature values(action) as action
values(file_name) as file_name values(file_hash) as file_hash count
earliest(_time) as first_event latest(_time) as last_event dc(index)
as cardinality_edr by dest
```

```
| where cardinality_edr = 2
| convert ctime(fist_event) ctime(last_event)
|  table  first_event  last_event  dest  signature  action  file_name
file_hash count
```

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
40	10	0	30	80

Tabla 77. Prioridad de la alerta UC02-AV-006

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
malware, amp, attack, alert, exploitation	critical, multi-av

Tabla 78. Etiquetado de la alerta UC02-AV-006

Aclara un hecho necesario para que funcione correctamente la alerta, la planificación debe poseer una ventana de tiempo lo suficientemente significativa como para que se puedan dar ambas correlaciones. A pesar de que es una alerta crítica, y por tanto debería de ejecutarse cada minuto, se ha planificado para que lo haga cada hora debido a que pueden o bien no correlacionarse correctamente la alerta debido a que los eventos surjan espaciados en el tiempo (léase una on-demand scan), o que por la propia naturaleza de la gestión del reenvío de eventos del EDR al SIEM, no se haga en un relativo real time, sino que el push se haga cada cierto tiempo.

3.13.4. Alertas Firewall.

En esta sección no nos vamos a centrar en la generación de alertas, que por otro lado daría para poder hacer decenas de éstas. Lo que vamos a hacer es mostrar una de las grandes ventajas y a su vez, funcionalidades que nos da el trabajar con Splunk. Y esta no es otra que las Apps que posee, y su interrelación con otros sistemas de seguridad más allá del SIEM.

En este ejemplo, vamos a integrar la plataforma de inteligencia de amenazas MISP con Splunk, y a su vez, vamos a realizar consultas complejas, relacionando los eventos notables de tipo los Firewall, con los valores contenidos dentro del MISP.

Integración del MISP con Splunk:

A continuación, se va a ver cómo se integra el MISP con Splunk. No se va a entrar demasiado en detalle, pero sí se va a dejar los pasos a dar para conseguir reproducirse.

Lo primero que vamos a hacer es descarga e instalar la APP. Para descargarla podemos hacerlo de dos formas. O bien, la podemos descargar desde la siguiente url, y luego subirla a Splunk; o bien lo podemos hacer directamente desde Splunk siguiendo los siguientes pasos:

Nos dirigimos a Find More Apps.

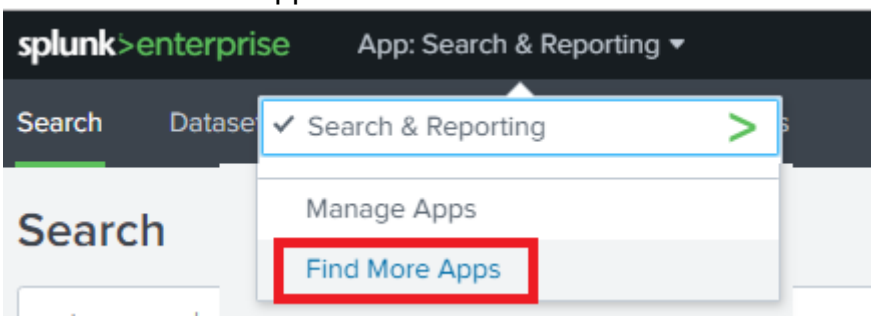

Ilustración 46. Búsqueda de la App

Luego buescamos por el nombre la APP que queremos instalar:

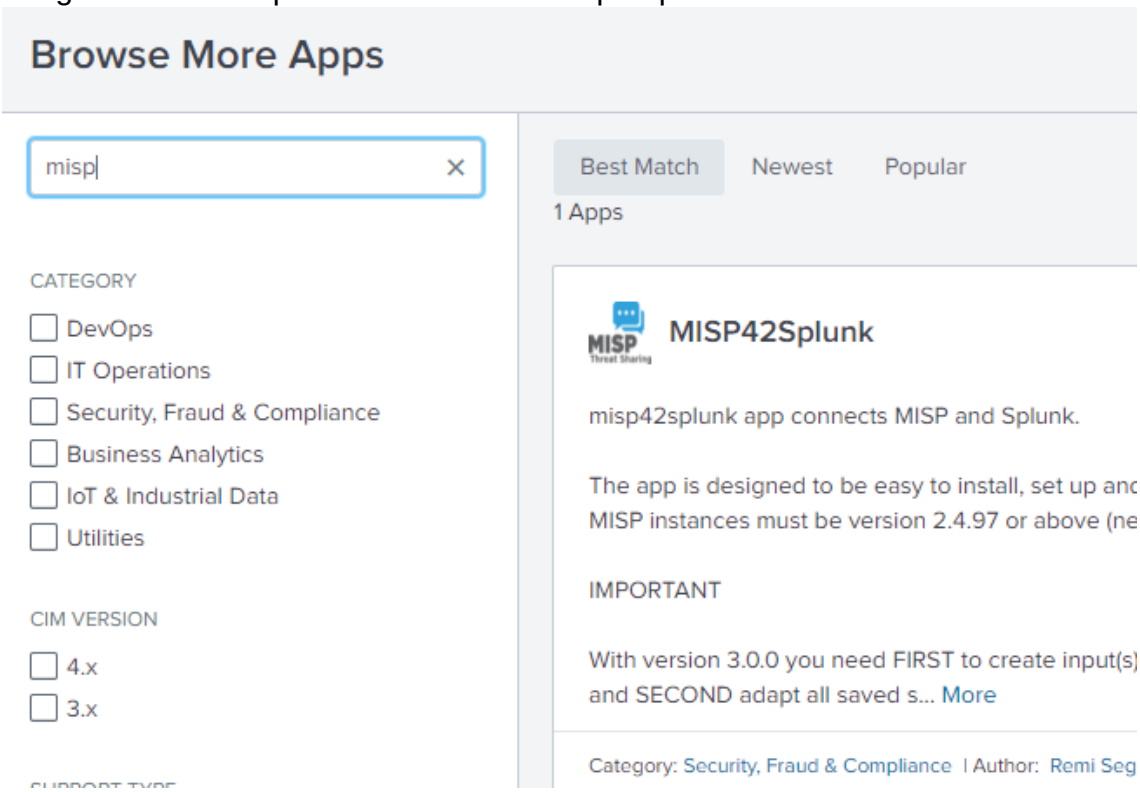

Ilustración 47. Navegamos para buscar la App

Una vez encontrada la APP, la instalamos a través del botón Install. Indicar, que debemos tener una cuenta de Splunk para poder instalar la APP (crearse una cuenta es gratuita).

Una vez instalada, podemos acceder a ésta, por ejemplo de la siguiente manera:

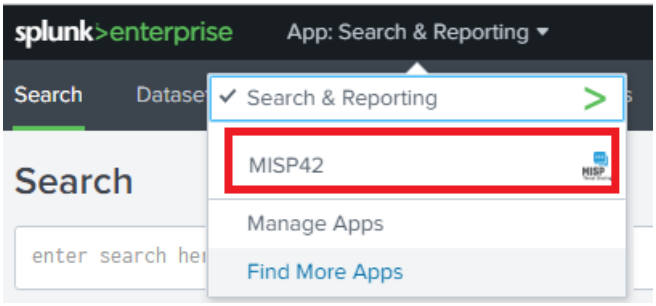

Ilustración 48. Podemos ver la APP instalada

Ahora nos dirigimos a la pestaña de Inputs para configurarla. Para ello pulsamos en el botón Add misp.

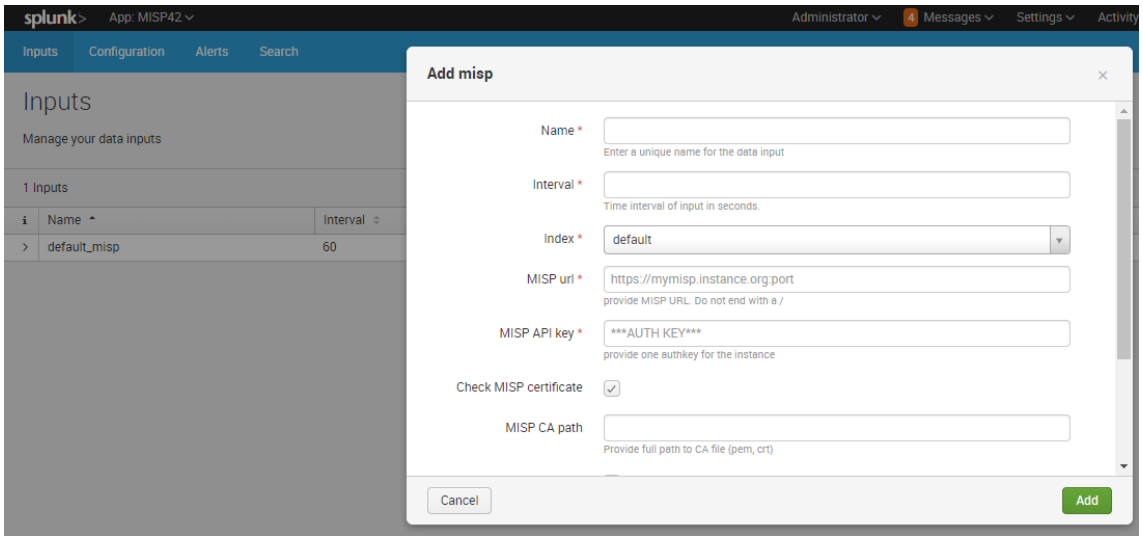

Ilustración 49. Parte de la configuración de la App

Los datos que se nos piden son los necesarios para conectarse a través de la API Rest del MISP. Se acaba por tanto de indicar la forma en que vamos a interactuar con el MISP. En esta APP ni se indexan los datos, ni se crean KV Store para el almacenamiento como ocurren en otras. La forma de interconectar por tanto es a través de consultas contra la API desde Splunk.

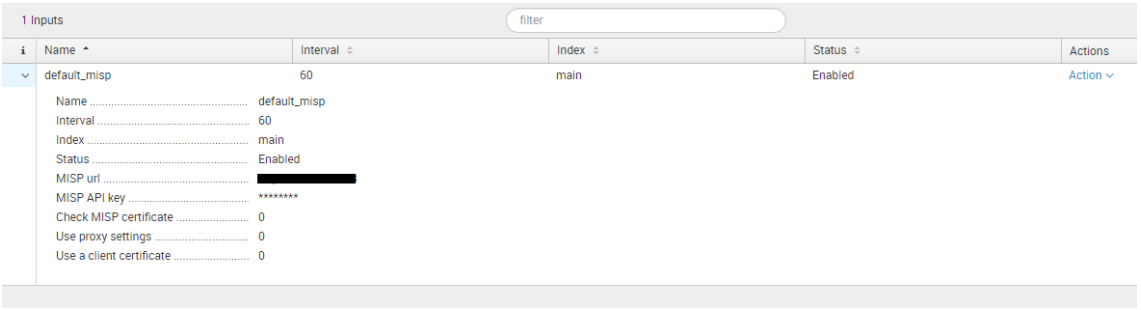

Ilustración 50. App configurada

Por último, se muestra una captura de pantalla con información básica de su funcionamiento.

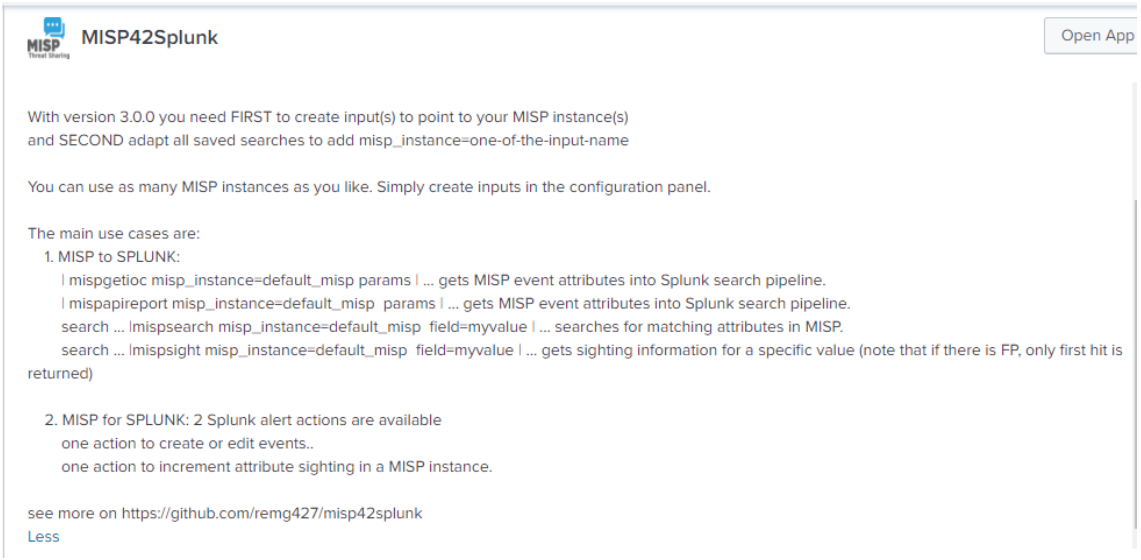

Ilustración 51. Información sobre el funcionamiento de la App

Para más información, podemos ver la documentación de la APP.

Ahora que lo tenemos integrado, podemos realizar múltiples alertas. Veamos un ejemplo.

UC03-FW-001 Detect IoC Firewall Traffic Outbound MISP

Nombre de la alerta	Descripción de la alerta
UC03-FW-001 detect IoC Firewall Traffic Outbound MISP	La presente alerta, lo que hace es en base a los eventos notables del Firewall, es buscar IP's de entrada que esté listada dentro del MISP.

Tabla 79. Descripción de la alerta UC03-FW-001

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC03-FW-001	Media	15 minutos	4 horas

Tabla 80. Información de la alerta UC03-FW-001

Veamos a continuación el código dentro de Splunk de la alerta.

```
index = main sourcetype = _json alerta = EN_Firewall_00*
     [  mispgetioc  misp_instance=default_misp  last=30d  type="ip-src"
to_ids=true warning_list=true limit=100000
    | rename misp_value as src_ip
    | table src_ip
    | format maxresults=60000]
|  stats  earliest(_time)  as  first_event  latest(_time)  as  last_event
values(rule) as rule values(dest_ip) as dest_ip values(vendor_action)
as vendor_action  dc(dest_ip) as cardinality_destip count by src_ip
| where cardinality_destip < 3
| eval dest_ip = mvindex(dest_ip, 0, 10)
| convert ctime(first_event) ctime(last_event)
| mispsearch misp_instance=default_misp field=src_ip
```

Como se puede ver, se ataca el MISP para obtener el tipo que nos interesa. Cualquier duda sobre el MISP la podemos consultar dentro de su documentación.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
35/40	10	0	10	57,5

Tabla 81. Prioridad de la alerta UC03-FW-001

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
authentication, windows, security, os, exploitation	critical, admin, success, access

Tabla 82. Etiquetado de la alerta UC03-FW-001

3.13.5. Alertas Cisco ESA.

Nos quedaría por ver algunas alertas referentes a los eventos notables del correo. Pero lo que vamos a hacer es mostrar es lo fácil que es relacionar dos fuentes distintas dentro de una consulta, cuando los campos están normalizados.

Vamos a suponer que tenemos un evento notable dentro del ESA que nos proporcionan las IP's desde donde se nos están enviando los correos. En un principio podríamos pensar que podrían ocupar mucho espacio en disco obtenerlas todas, pero tenemos la ventaja (de las pocas a la hora de tratar esta fuente) que los logs del ESA se indexan línea a línea separados. Esto hace que

si bien su reconstrucción sea compleja, el evento que nos interesa en realidad ocupa muy poco espacio.

Por ejemplo, a continuación podemos ver los logs de una conexión entrante bloqueada. Cada línea es un log distinto enviado al Data Lake:

```
2019-12-08T09:15:17.000Z xxxxxxx.local <22>Dec 08 10:15:17 - 10.164.26.30: Info: New SMTP ICID 4199074 interface Perimetral (192.168.100.17) address 90.189.110.166 reverse dns host sng.khakasnet.ru verified yes

2019-12-08T09:15:17.584Z xxxxxxx.local <22>Dec 08 10:15:17 - 10.164.26.30: Info: ICID 4199074 REJECT SG BLACKLIST match sbrs[-7.0:-3.0] SBRS -4.0 country Russian Federation

2019-12-08T09:15:17.584Z xxxxxxx.local <22>Dec 08 10:15:17 - 10.164.26.30: Info: ICID 4199074 close
```

Por tanto, podemos enviarlos como eventos notables y correlacionarlos. Supongamos que a dicho evento notable lo hemos denominado EN_ESA_003. Como campos contiene únicamente la fecha del evento, la IP y el dominio. Simplemente con esta información nos bastará para realizar la correlación que queremos.

Como habíamos visto antes, tenemos un evento notable de tipo Firewall (EN_Firewall_001) que nos muestran posibles IP's maliciosas. Por tanto, podemos detectar posibles conexiones maliciosas contra las pasarelas de correos (básicamente envío de correos maliciosos).

Dado que el objetivo final de la alerta es el detectar posibles correos maliciosos, dentro de la nomenclatura de la alerta, usaremos la dedicada a las alertas de correo. En otras ocasiones tendremos alertas que realmente será complicado determinar un objetivo tan claro. Para estos casos se utilizarán nomenclaturas específicas.

UC04-MAIL-001 Suspicious mail connection

Nombre de la alerta	Descripción de la alerta
UC04-MAIL-001 Suspicious mail connection	Alerta que busca detectar conexiones entrantes contra la infraestructura de correo.

Tabla 83. Descripción de la alerta UC04-MAIL-001

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC04-MAIL-001	Baja	Cada 30 minutos	8 horas

Tabla 84. Información de la alerta UC04-MAIL-001

La alerta correlada es bastante simple como se puede ver a continuación:

```
index = main sourcetype = _json ((alerta = EN_Firewall_001 AND
dest_port = 25) OR alerta = EN_ESA_003)
| stats dc(alerta) as cardinality_alert earliest(_time) as first_event
latest(_time) as last_event count by src_ip
| where Cardinality_alert = 2
```

Para el cálculo de la prioridad no vamos a tener en cuenta en este caso que pase por dos appliance de seguridad, dado que es obvio que si la arquitectura de red de la empresa está bien hecha, cualquier conexión entrante a la pasarela de correo debe, en primer lugar, pasar por un firewall.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	0	0	0	20

Tabla 85. Prioridad de la alerta UC04-MAIL-001

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
attack, email, incoming, network, delivery	low, suspicious

Tabla 86. Etiquetado de la alerta UC04-MAIL-001

Con esta simple alerta, hemos logrado detectar una IP sospechosa dentro del appliance de seguridad de correo. Posteriormente se analizará la acción del propio ESA, pero eso ya será labor del analista, o bien dependiendo del volumen de esta alerta, se podría realizar correlaciones más complejas, debido a que podríamos ver qué IP's son bloqueadas por el ESA, y descartarlas de dicha alerta. De esta forma tendríamos una alerta más elaborada.

Para lo anterior deberíamos tener un evento notable dentro de los Cisco ESA, que nos enviara todas las IP's que son detectadas como maliciosas por su motor de reputación (EN_ESA_004). Veamos cómo quedaría dicha alerta:

UC04-MAIL-002 Suspicious mail connection non-blocked

Nombre de la alerta	Descripción de la alerta
UC04-MAIL-002 Suspicious mail connection non-blocked	Alerta que busca detectar conexiones entrantes contra la infraestructura de correo, y que el Cisco ESA no la ha detectado como maliciosa a nivel de reputación.

Tabla 87. Descripción de la alerta UC04-MAIL-002

Identificador alerta	prioridad	Planificación temporal	silenciamiento
UC04-MAIL-002	Media	Cada 15 minutos	4 horas

Tabla 88. Información de la alerta UC04-MAIL-002

La correlación dentro de Splunk es la siguiente:

```
index = main sourcetype = _json ((alerta = EN_Firewall_001 AND
dest_port = 25) OR alerta = EN_ESA_004)
| stats values(alerta) as alertas dc(alerta) as cardinality_alert
earliest(_time) as first_event latest(_time) as last_event count by
src_ip
| where Cardinality_alert = 1 AND match(alertas, "EN_Firewall_001"
```

La correlación es bien sencilla, si la IP sospechosa va hacia el puerto 25, y no está dentro de las IP's del evento notable 004, esto querrá decir que no ha sido detectada por el ESA como una IP maliciosa.

Cálculo de la prioridad:

Prioridad de base	# targets > 1	Mon. crítica	Factor corrección	Prioridad final
20	0	0	20	40

Tabla 89. Prioridad de la alerta UC04-MAIL-002

Etiquetado:

Etiquetas del Evento Notable	Nuevas etiquetas
attack, email, incoming, network, delivery	medium, suspicious, allowed

Tabla 90. Etiquetado de la alerta UC04-MAIL-002

Aun así, esto no quiere decir que el Cisco ESA posteriormente no haya hecho alguna acción de bloqueo. Deberá ser analizado por el equipo de seguridad.

Y como estas alertas, podemos realizar muchas más. Por ejemplo, como tenemos integrado el MISP con nuestro SIEM, podemos realizar correlaciones entre los eventos del ESA y los IOCs contenidos dentro de esta plataforma.

3.13.6. Resto de alertas.

Hasta aquí hemos visto un pequeño ejemplo de las cosas que podemos hacer usando eventos notables, la normalización de los campos y un la complejidad de un SIEM.

Tanto los eventos notables, como sus posteriores alertas dependen obviamente de la propia infraestructura con la que estemos trabajando. De ésta dependerán los tipos de appliances de seguridad que dispongamos para analizar eventos, de los sistemas que protejamos y sus criticidades, … por ello es difícil crear un librería de eventos notables y alertas estándar.

Sí es cierto que se pueden implementar alertas conceptualmente independientes del dispositivo como hemos podido ver con las alertas del AV, pero a mayor detalle de los eventos reportados por éste, y de sus funcionalidades, mayor número y mejores correlaciones, detecciones, … podremos realizar dentro del SIEM.

La elaboración por tanto de este aspecto del SIEM dependerá mucho del equipo de personas que se encarguen de realizarlas y de sus conocimientos. Es relativamente fácil el desplegar unas cuantas alertas genéricas, lo complicado es poder afinar sus parámetros para adaptarlos a cada circunstancia y organización. Si se posee un equipo altamente especializado en análisis de amenazas y Threat Hunting, se pueden transformar las consultas que se suelen usar, en alertas, reportes y Dashboards dedicados. De esta forma, se entra en un estado de madurez alto en cuanto a la detección de amenazas, y la gestión de incidentes de seguridad.

En el presente trabajo no se va a desarrollar más dicha base de datos de alertas y eventos notables, dado que no es objetivo de éste. Sí indicar, que si se desea poseer alertas genéricas se pueden llegar a implementar de forma relativamente rápida y sencilla, existe bastante documentación (eso sí, dispersa) dentro de internet, como para poder crearlos. Lo que debemos entender que es aquí precisamente, dentro de la inteligencia de amenazas, donde un SIEM adquiere todo su poder. Por ello, todos los SIEM comerciales no hacen públicas sus alertas, dado que son su gran valor. Es más, en el ámbito profesional, nos vamos a encontrar lo mismo con los MSSP y sus clientes. Por tanto, la importancia de tener dentro del SOC a un gran equipo de inteligencia que desarrolle estas alertas, adaptándose al entorno es fundamental.

Por último indicar, que tan importantes son las alertas desplegadas como tener una amplia gama de Dashboards y reportes para la posterior investigación (sin olvidar el saber realizar consultas dentro de éste). La idea de madurez dentro de un SIEM y un SOC es que, la investigación se realice casi en exclusiva dentro del SIEM, y que no haga falta acceder a los appliances de seguridad o las fuentes que envían los eventos para su análisis. Sí es cierto, que para obviamente, en un DFIR vamos a tener que salir del SIEM para poder realizarlo. Lo mismo pasará con el análisis de malware. Pero ya estamos tratando temas como son la respuesta ante incidentes e inteligencia dentro del SOC, que son realizados por equipos especializados. La realidad es que, las principales funciones de un SOC pasarán por la relación de sus integrantes con el SIEM.

3.13.7. Otros aspectos a destacar en la implementación.

Lo que debemos hacer una vez que tenemos definidos los distintos eventos notables, alertas y están los campos normalizados, es definir las distintas identidades y assets críticos que están dentro de nuestra infraestructura.

Assets críticos.

Para la definición de los assets críticos, normalmente se suele usar la información contenida dentro de una CMDB. En caso de no tenerla, podemos crear tablas con dicha información. Esto va a ser fundamental a la hora de tratar las alertas, dado que su prioridad y severidad cambiarán en función de si estamos tratando un posible incidente dentro de uno de estos sistemas o usuarios.

Una forma sencilla de implementar esto en Splunk, es con la creación de un CSV que contenga toda esta información. Un ejemplo son los usuarios administradores del dominio dentro las máquinas Windows.

Podemos crear una tabla en formato CSV dentro de Splunk a nivel de interfaz web, de las siguientes dos formas:

La primera es mediante la instalación y el uso de una APP de Splunk llamada Lookup File Editor.

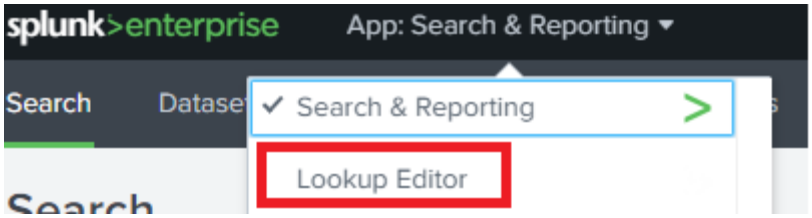

Ilustración 52. Accedemos a la App

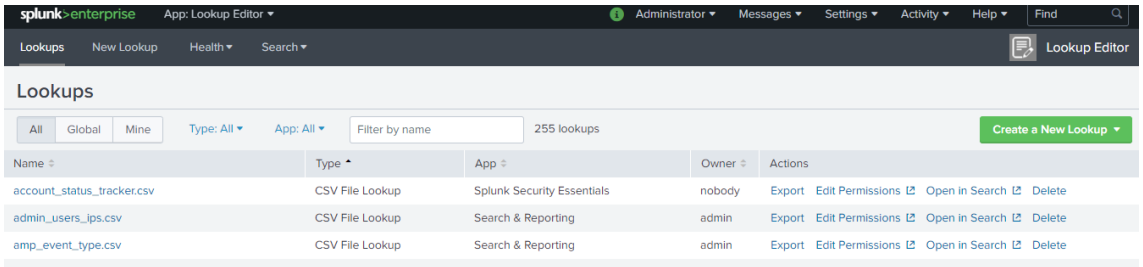

Ilustración 53. Listamos los CSV que existen actualmente

Donde seleccionamos el tipo de Lookup que queremos crear. Para el ejemplo, crearemos un CSV lookup.

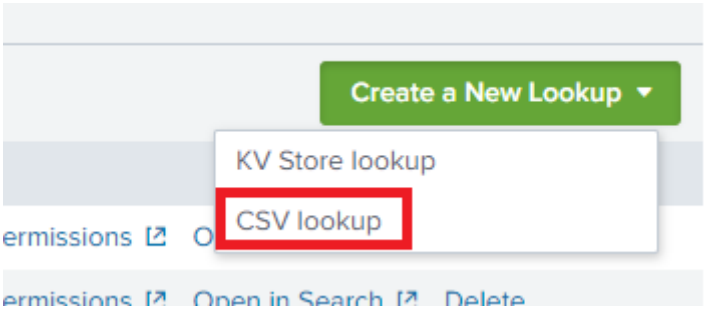

Ilustración 54. Creamos un nuevo CSV

A continuación se nos mostrará el editor del CSV.

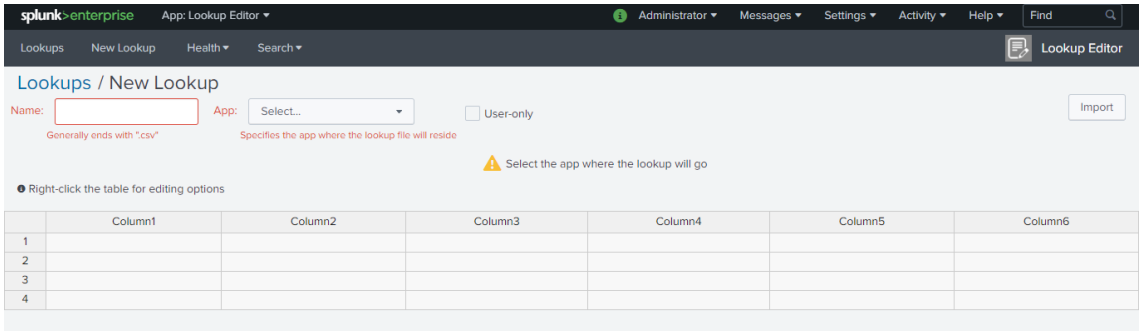

Ilustración 55. Página para la edición del CSV

Añadiremos los valores de la siguiente manera:

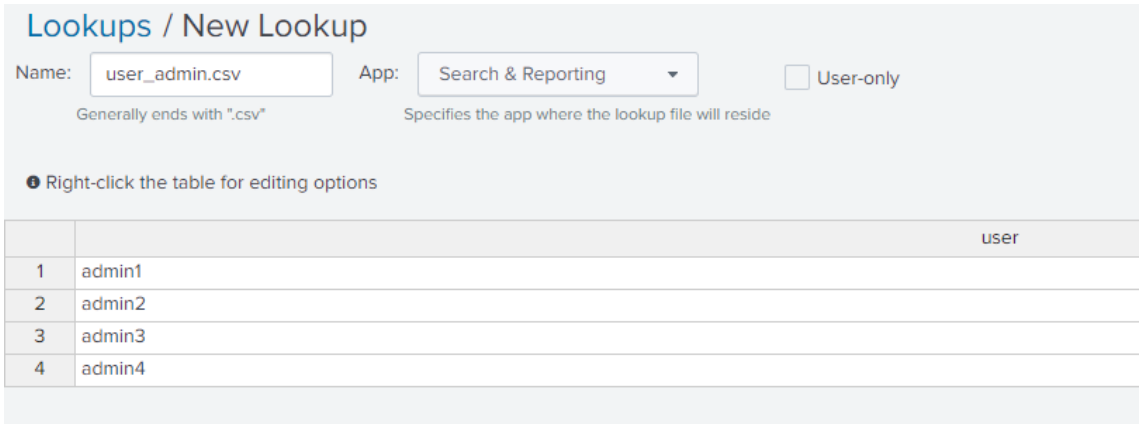

Ilustración 56. Se añaden la información

Guardándose una vez que tenemos los valores deseados.

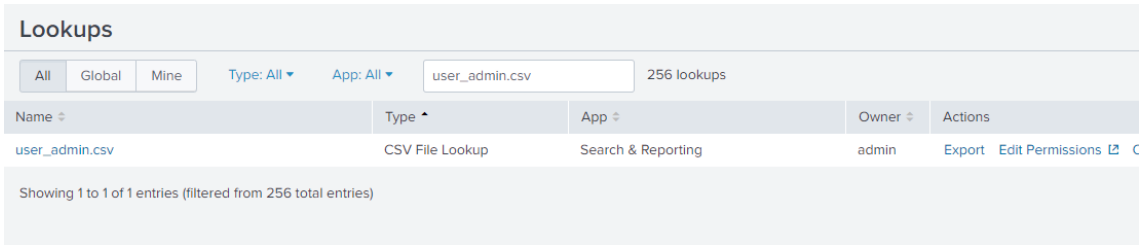

Ilustración 57. Guardamos la información

La segunda forma permite subir un fichero ya creado en formato CSV:

Dentro de Settings, accedemos dentro de los objetos de conocimiento a la opción Lookups.

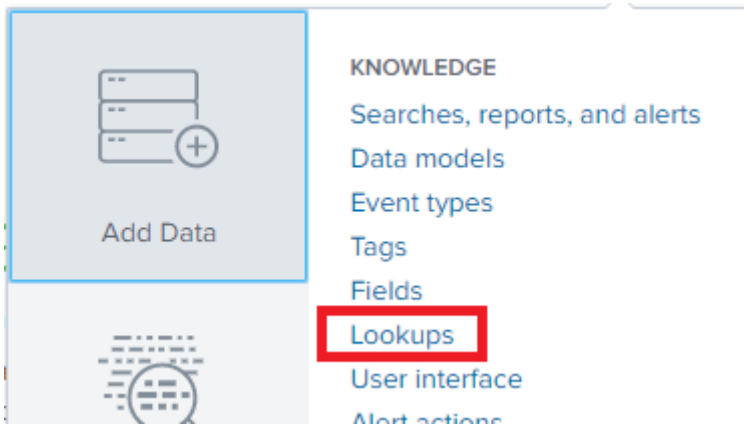

Ilustración 58. Otra forma de crear un CSV

Le indicamos que queremos crear un nuevo Lookup table files.

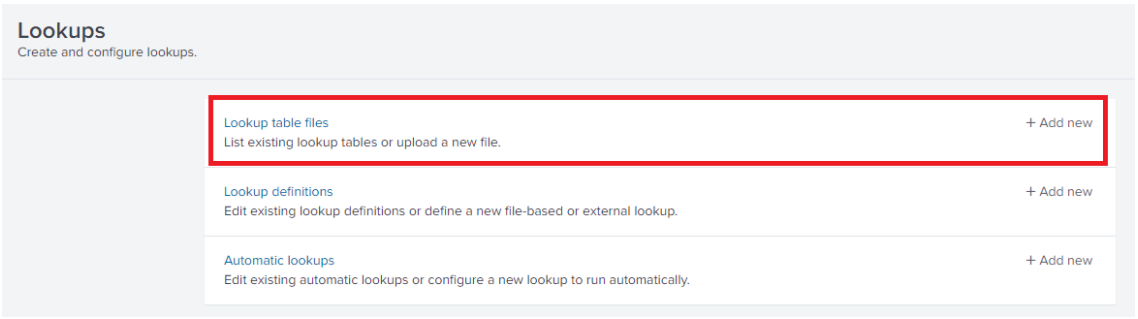

Ilustración 59. Se añade el CSV

Y subimos el fichero que queramos que se guarde.

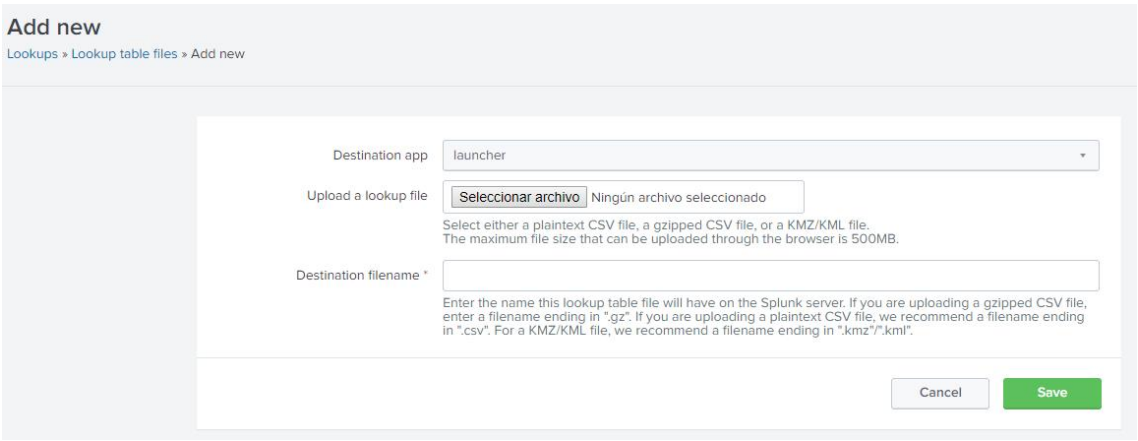

Ilustración 60. Modo de subir el CSV

La forma de usarlo dentro de Splunk es muy sencillo. Lo podemos usar de dos formas distintas. Podemos tener una Add-on que haga la importación del CSV, compruebe si el usuario está dentro de la lista, y si lo está, crear un nuevo campo con un valor identificativo de lo anterior.

La segunda forma, y la que vamos a implementar en este ejemplo, es realizar lo anterior dentro de la propia consulta. Veámoslo con un evento notable de Windows de tipo 4625.

```
index = main sourcetype = _json alerta = EN_Windows_4625
| search [ | inputlookup user_admin.csv ]
| stats count by user
| where count > 50
```

La alerta anterior, correla de forma básica eventos de tipo 4625, cuyo usuario esté dentro de la lista de usuarios administradores, y que en la ventana de tiempo indicada por la alerta, haya más de 50 eventos.

Como se ve, la forma de utilizar los CSV es en extremo sencilla y potente.

Indicar, que los SIEM de última generación son capaces de integrarse con el Directorio Activo para obtener la información que hemos creado anteriormente.

Gestión de identidades.

Otro de los elementos clave a la hora de implementar un SIEM maduro es ser capaz de detectar eventos sospechosos de una misma identidad digital. En este caso por identidad digital vamos a entender todos los campos normalizados que pueden identificar a un usuario o máquina dentro de la red.

Por ejemplo, un usuario puede ser identificado tanto por su nombre de usuario dentro de la organización, por su correo electrónico, por su IP y nombre de máquina en caso de tener una asignada de forma personal, … Toda esta información puede estar tanto dentro de una CMDB, como dentro de una lista personalizada creada ac-hoc, o ambas a la vez.

Veamos la forma de implementar lo anterior:

La siguiente aproximación es solo orientativa. La idea es poder realizar consultas contra todos los logs indexados, buscando agruparlos posteriormente por las distintas identidades existentes.

Como ejemplo sencillo, se crea un CSV con la información que los campos principales identificativos de un usuario determinado. Hemos supuesto en este caso, que un usuario de la red puede ser identificado mediante cuatro campos distintos: nombre de usuario, IP de la máquina habitual, nombre del host habitual y la cuenta de correo.

El CSV se creará de la forma que hemos visto anteriormente. En este caso se le ha llamado identidades.csv Dicha tabla posee las siguientes columnas:

- ident. Se trata de un número que identifica a cada persona).
- user. Campo con el nombre de usuario.
- ip. Campo que contiene la IP del Workstation habitual del usuario.
- computername. Campo que recoge el nombre del Workstation habitualmente usado por dicho usuario.
- mail. Nombre de la cuenta del usuario de correo corporativo.

Un código en Splunk válido sería el siguiente:

```
index = main sourcetype = _json (alerta = * OR source = UC*)
| lookup indentidades.csv ip as src_ip output ident as id1
| lookup indentidades.csv ip as dest_ip output ident as id2
| lookup indentidades.csv computername as host output ident as id3
| lookup indentidades.csv user output ident as id4
| lookup indentidades.csv mail as src_user output ident as id5
| lookup indentidades.csv mail as recipient output ident as id6
| table alerta, source, id1, id2, id3, id4, id5, id6, src_ip, dest_ip,
host, user, src_user, recipient
|  where isnotnull(id1) OR isnotnull(id2) OR isnotnull(id3) OR
isnotnull(id4) OR isnotnull(id5) OR isnotnull(id6)
| eval identity = if(isnotnull(id1), id1, if(isnotnull(id2), id2,
if(isnotnull(id3), id3, if(isnotnull(d4), d4, if(isnotnull(d5), d5,
d6)))))
| stats values(source) as source dc(source) as distinct_sources
values(alerta) as alerta dc(alerta) as distinct_alerts count by
identity
```

La idea es simple, buscamos en cada alerta o evento notable indexado, uno de los 4 campos incluidos dentro del CSV. Indicar, que la IP puede ser tanto de origen, como de destino, y que el correo electrónico también puede ser tanto de entrada como de salida. Por tanto, nos queda un total de 6 posibles identificadores distintos. En cada uno de estos eventos, lo que hacemos es crear un nuevo campo con el valor numérico dado al usuario.

Una vez que hemos creado o no dichos nuevos campos, se filtran todos los eventos en los que existan estos. Por último, se hace la agrupación por el campo identity, mostrando a su vez el número total de eventos y las alertas o eventos notables a los que corresponden.

Esto es solo una pequeña aproximación de lo que se puede hacer, se le podría ir pasando más campos como pueden ser las etiquetas, y crear campos valuados para darles pesos a dichos resultados y así poder crear nuevas alertas.

Otros aspectos relevantes.

Antes de poner una alerta en producción se deben realizar las siguientes comprobaciones:

- Análisis de los distintos umbrales de las alertas. Actualmente, los SIEM de tercera generación se basan en el Machine Learning para aprender y adaptar estos umbrales en función de los propios eventos contenidos en estos. Esto hace que se reduzcan drásticamente los falsos positivos y la fatiga del analista por exceso de alertas. Dado que en nuestra aproximación no vamos a disponer de estas funcionalidades, debemos mantener las alertas que vayamos creando en un estado de preproducción durante un tiempo lo suficientemente largo para poder detectar las mayores anomalías posibles que no supongan una amenaza real, para de esta forma determinar correctamente dichos umbrales.

- Número de alertas que se disparan. Como se ha comentado anteriormente, antes de poner en producción una alerta, debemos determinar si la alerta va a generar demasiado ruido. En esto también influirá la prioridad de la alerta. Es posible, que una alerta genere excesivo ruido con una baja prioridad. En este caso, lo que haremos antes de ponerla en producción es pasarla a reporte. La diferencia es que en lugar de generarse un ticket por evento, se genera un ticket por disparo, independientemente del número de eventos que se hayan generado. La planificación temporal de los reportes aumenta con respecto a las alertas (para que tenga realmente un efecto de reporte, y no de alerta a nivel de número de tickets generados). Por tanto, deberemos tener esto en cuenta a la hora de pasar una alerta a reporte.

Gestión de las etiquetas.

Las etiquetas no son únicamente elementos informativos. También nos pueden servir para detectar debilidades a nivel de implementación de los eventos notables y de las alertas, e incluso a nivel de arquitectura de seguridad.

Si agrupamos los eventos notables o las alertas por el campo del Cyber Kill Chain, podemos ver tendencias de ataques, debilidades dentro de los appliances de seguridad, o incluso de las políticas de seguridad implementadas por la organización.

Por supuesto, estos valores se gestionan a través de los KPIs y KRIs.

A nivel de etiquetado de las alertas, hemos creado etiquetas custom, y basadas en el CKC. Pero a medida que conseguimos madurar nuestros EN y alertas, podemos ir añadiendo etiquetas de tipo Mitre ATT&CK y CIS. Estas etiquetas

nos ayudarán para la realización de Threat Hunting, y la detección tanto de APTs, como de su actores implicados.

Una vez que tengamos implementados estas etiquetas en base a los EN y alertas, seremos capaces de crear Dashboards y otras nuevas alertas para detectar lo anterior.

3.13.8. Tabla resumen de las alertas implementadas.

Identificador alerta	Nombre de la alerta	Prioridad
UC01-WIN-001	Multiple Failed Logins from different accounts same host	Media – 55
UC01-WIN-002	Multiple Failed Logins from same account same host	Media – 55
UC01-WIN-003	Multiple failed logins of non-existing accounts same host	Media – 55
UC01-WIN-004	Multiple failed logins of non-existing account different hosts	Alta – 65
UC01-WIN-005	Suspicious attempt to logon with Guest account	Alta – 65
UC01-WIN-006	Suspicious Failed Logon error codes	Media – 55
UC01-WIN-007	Multiple brute-force authentication attempts in different hosts	Alta – 65
UC01-WIN-008	Logon attempt with Guest account – Error Code 0xC000006A	Alta – 65
UC01-WIN-009	Multiple failed login attempts successful login with Guest account	Crítica – 80
UC01-WIN-010	Multiple failed login attempts successful login with admin account	Crítica – 80
UC02-AV-001	Several detection same host	Alta – 60
UC02-AV-002	Several detection same malware multiple hosts	Crítica - 80
UC02-AV-003	Several detections same malware same host	Alta – 60
UC02-AV-004	Virus detected and not removed	Crítica – 90
UC02-AV-005	AV Report	Alta – 60

UC02-AV-006	Threat detect multi AV	Crítica – 80
UC03-FW-001	Detect IoC Firewall Traffic Outbound MISP	Media – 57,5
UC04-MAIL-001	Suspicious mail connection	Baja – 20
UC04-MAIL-002	Suspicious mail connection non-blocked	Media – 40

Tabla 91. Alertas resumen existentes

3.14. Generación Dashboard para gestión y monitorización de las alertas y eventos notables.

En esta sección, vamos a ver una de las funciones más importantes de un SIEM. Nos estamos refiriendo a la implementación de Dashboards.

Indicar que, los Dashboards principales de análisis de eventos los tenemos desplegados dentro del Data Lake (en concreto, dentro del Kibana). Por tanto, para el análisis de los posibles incidentes, o análisis forense, junto con las políticas de retención, están desplegadas dentro de Elastic.

Únicamente, y de momento, dentro de Splunk se van a implementar un único Dashboard. Este Dashboard mostrará información estadística sobre los eventos notables y alertas que están generándose. Se mostrará tanto información del número de eventos que se producen, como información sobre las etiquetas creadas.

Veamos pues, el Dashboard que se ha implementado:

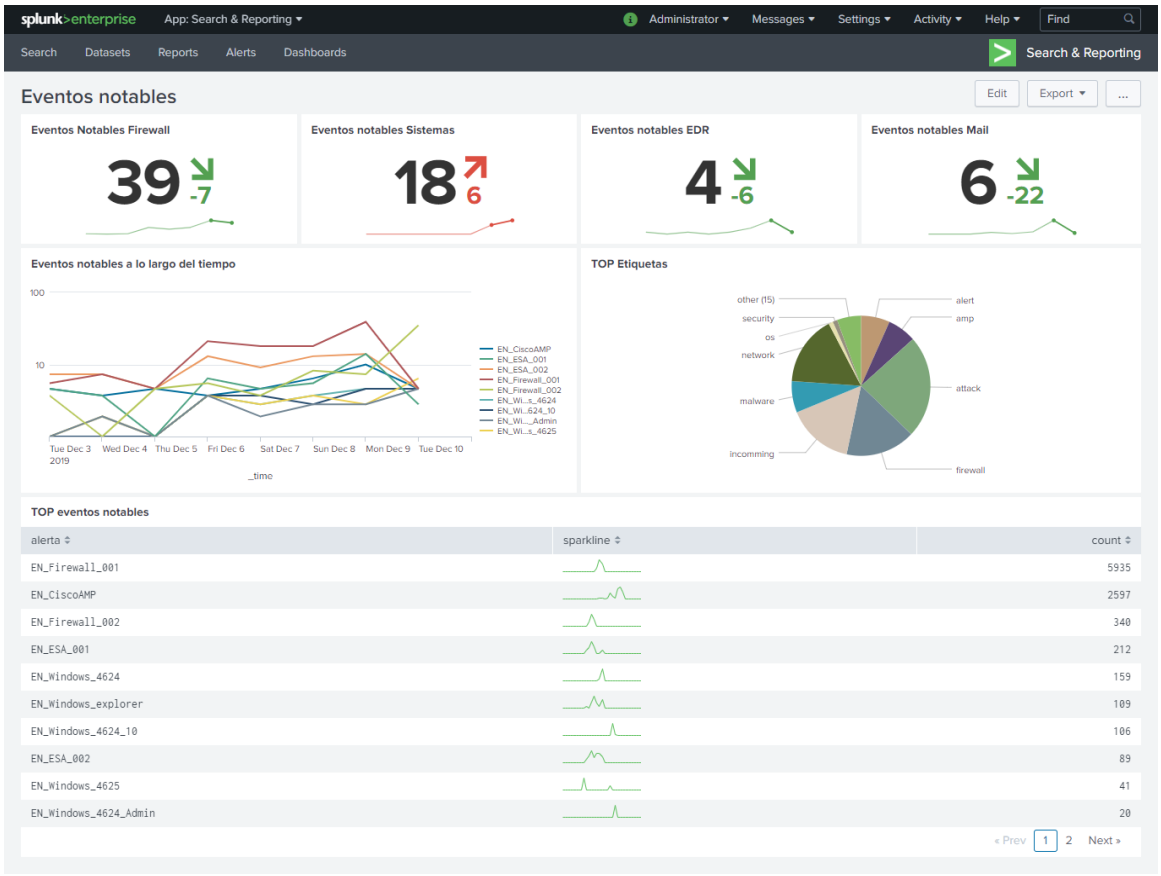

Ilustración 61. Dashboard Eventos Notables

Es simplemente una aproximación de las posibilidades que se puede llegar a tener. Se ha realizado basándose únicamente en los eventos notables, se podría añadir nuevos paneles referentes a las alertas.

Este Dashboard nos sirve por tanto a alto nivel el poder ver estadísticas de los distintos eventos notables que se van disparando.

Veamos los distintos bloques que contienen el Dashboard.

En los siguientes paneles relacionados entre sí, podemos ver el número de eventos notables por familia que han saltado en el día de hoy (número más grande), con respecto a la diferencia en el día anterior. Interpretamos por tanto el primer panel como que en el día de hoy se han generado un total de 39 eventos notables de tipo Firewall, lo cual son 7 eventos notables menos que en el día anterior. Se marca el descenso en verde, dado que esto en sí es positivo.

Ilustración 62. Paneles EN por categorías

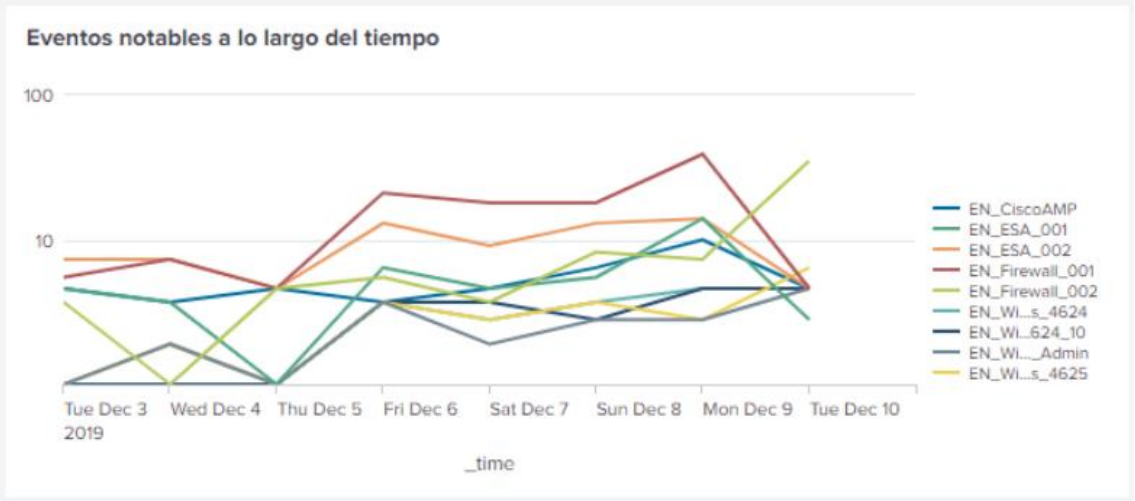

Ilustración 63. EN a lo largo del tiempo

En la siguiente gráfica podemos ver un gráfico de eventos notables y su evolución a lo largo de los últimos 7 días. En el eje de las ordenadas podemos ver el número total de eventos que se han producido, mientrás que en el eje de las ascisas vemos las marcas de tiempo.

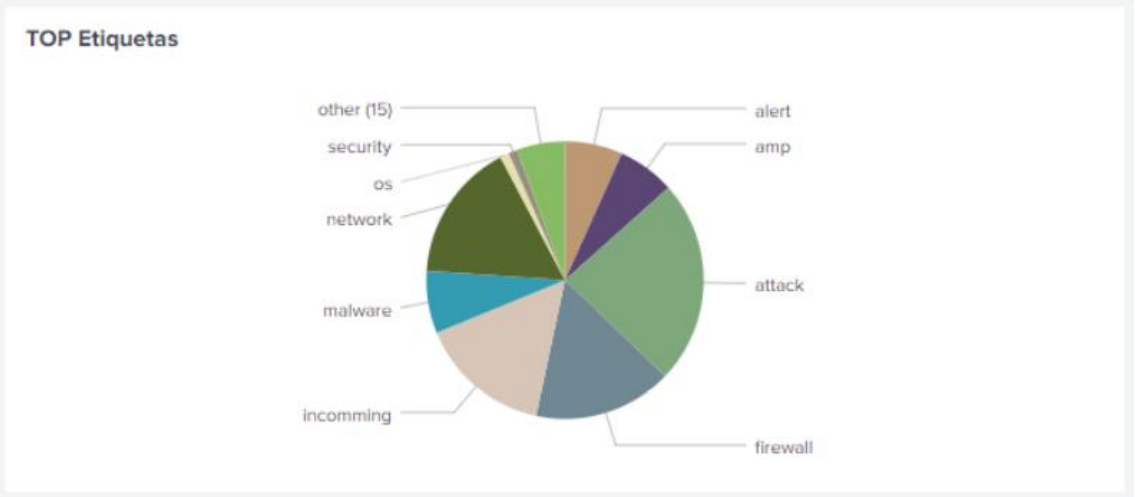

Ilustración 64. TOP de etiquetas de los eventos notables

En el presente gráfico circular, podemos ver la proporción de las distintas etiquetas que poseen los eventos notables. Podemos ver que la etiqueta attack en este caso es la que más veces ha salido en el total de los eventos.

Ilustración 65. Eventos notables y su volumetría

Por último, en el panel inferior podemos ver el top de eventos notables ordenados por el número total de ellos. De esta forma podemos hacernos una idea de los eventos notables que más saltan.

Como se puede ver y se ha comentado anteriormente, se trata de un Dashboard de alto nivel. Podríamos haber creado más paneles para ver otras caracterísitcas como pueden ser el agrupamiento de los eventos notables por prioridades, o similar.

En todo caso, para poder ver los eventos en detalle podemos abrir una ventana de búsqueda (Search) y realizar ahí a través del SPL, consultas desde las más simples a las más complejas. Indicar, que si determinamos que una consulta puede ser interesante para el Dasborad, ésta se puede guardar como panel de un Dashboard a través del botón Save As > Dashboard Panel. El formato resultado de la búsqueda (ya sean un formato tabla, charts, Single value, …).

Para consultar información referente a la creación de Dashboards y visualizaciones dentro de Splunk, podemos consultar la siguiente documentación.

3.15. Creación de los Playbooks para respuesta ante incidentes en base a las alertas SIEM.

Toda alerta creada debe ser gestionada por los integrantes del equipo de seguridad. La forma de hacerlo puede variar dependiendo del modelo de SOC que se tenga. Para el presente trabajo, vamos a suponer que existen tres niveles (L1, L2 y L3), y que en la gestión de las alertas se va a poder comunicarse con otros departamentos implicados dentro de la organización.

Dado que no es objetivo principal del TFM el profundizar en la respuesta ante incidentes, únicamente se va a documentar alguna alerta de ejemplo. Se va a realizar el procedimiento de una de cada tipo de alertas por familia que se han creado.

Procedimientos.

Alerta UC01-WIN-010

Nombre de la alerta: UC01-WIN-010 Multiple failed login attempts successful login with admin account

Descripción: La alerta UC01-WIN-010 se encarga de detecta múltiples intentos fallidos de login con una cuenta de administrador, justo antes de un evento satisfactorio de acceso. La alerta lo que intenta detectar es un ataque exitoso contra contraseña (fuerza bruta, ataque diccionario, …) de una cuenta con privilegios de administrador de dominio.

Prioridad/severidad: Crítica

Procedimiento:

Pasos a realizar	Nivel
1. Dado que es un evento crítico, la alerta se escala directamente a L2.	L1
1. Analizar los eventos generador de la alerta, buscando detectar posible error en la correlación. a. En caso de determinarse que se trata de un error de correlación, cerraremos el incidente como Falso Positivo, y se procederá a su revisión y corrección. 2. En caso de confirmarse que la correlación es correcta. a. Contactar telefónicamente con el departamento de Sistemas para proceder al bloqueo del usuario. b. Contactar con el usuario para confirmar legitimidad o no de la acción descrita en la alerta. 3. En caso de que la acción haya sido realizada por el usuario administrador. a. Desbloquear la cuenta. b. Cerrar el presente incidente como Falso Positivo. 4. En caso de confirmarse el compromiso. a. Recopilar toda la información relacionado con el incidente (origen de la máquina desde donde se realiza la acción, cuenta comprometida, …). b. Escalar a L3 para activar procedimientos de	L2

cuenta comprometida, máquina comprometida, … Todo ello definido dentro de los Playbook de respuesta ante incidentes.	
1. Aplicar los procedimientos de actuación descritos dentro de los documentos para la respuesta ante incidentes.	L3

Tabla 92. Procedimiento UC01-WIN-010

Alerta UC02-AV-004

Nombre de la alerta: UC02-AV-004 Virus detected and not removed

Descripción: La alerta se encarga de detectar un evento generado por el EDR de una detección de amenaza no eliminada.

Prioridad/severidad: Crítica

Procedimiento:

Pasos a realizar	Nivel
1. Dado que es un evento crítico, la alerta se escala directamente a L2.	L1
1. Analizar a partir del evento generado, el motivo por el cuál no se ha podido eliminar la amenaza. a. En caso de detectar un motivo justificativo del error, cerrar el incidente como Sin Impacto. 2. Si se confirma el compromiso de la máquina, escalar a L3 para ejecutar el procedimiento de máquina comprometida.	L2
1. Ejecutar el procedimiento de máquina comprometida. En este procedimiento se ejecutan acciones de contención como el sacar al equipo de la red, y acciones relacionadas con el análisis forense.	L3

Tabla 93. Procedimiento UC02-AV-004

Alerta UC03-FW-001

Nombre de la alerta: UC03-FW-001 Detect IoC Firewall Traffic Outbound MISP

Descripción: La alerta se vale de los eventos notables generados por el firewall, confrontándolos con los eventos del MISP para detectar algún indicio de compromiso interno.

Prioridad/severidad: Media

Procedimiento:

Pasos a realizar	Nivel
1. Recopilar toda la información de la alerta. Especialmente toda la información referente a la IP de origen y la máquina o máquinas de destino. 2. Comprobar la reputación de la IP atacante. 3. Escalar toda la información a L2.	L1
1. Analizar dentro del MISP toda la información referente al IoC de la alerta. 2. En base a toda la información dada por el MISP, analizar si hay más coincidencias dentro de los eventos almacenados dentro del Data Lake. a. Si después del análisis, se determina que no está relacionado con la información contenida dentro del MISP, cerrar el incidente como Falso Positivo. 3. Si determinamos que el evento del firewall está relacionado con lo descrito dentro del MISP. a. Incrementar la severidad de la alerta a alta, y escalar a L3 para que se aplique el procedimiento de máquina comprometida. 4. Si el evento no está relacionado con lo detectado por el MISP, determinar si se trata de una acción sospechosa de otro tipo. a. En caso de no detectarse acción sospechosa, cerrar el incidente como Falso Positivo. b. Si se detecta posible compromiso, escalar a L3 para la aplicación del procedimiento de máquina comprometida.	L2
1. Aplicar el procedimiento de máquina comprometida.	L3

Tabla 94. Procedimiento UC03-FW-001

Alerta UC04-MAIL-001

Nombre de la alerta: UC04-MAIL-001 Suspicious mail connection

Descripción: La alerta relaciona eventos notables tanto de firewall (posible IP atacante), como del appliance de seguridad de correo, para determinar la posible entrada de un correo malicioso.

Prioridad/severidad: Baja

Procedimiento:

Pasos a realizar	Nivel
1. Recopilar la información relacionada con el incidente. 2. Comprobar la reputación de la IP de origen.	L1
1. Comprobar la acción realizada por el ESA. a. Si la conexión ha sido bloqueada o el correo ha sido bloqueado por el ESA y no se considera necesario realizar otra acción, cerramos el incidente como Sin Impacto. 2. En caso de que el correo haya sido reenviado por el ESA. Analizar el correo sospechoso. a. Si se determina que es sospechoso, escalar el incidente a L3 para su análisis forense, y la aplicación de medidas de respuesta ante incidentes. b. En caso de determinar que el correo es benigno, cerrar el incidente como Falso Positivo.	L2
1. Analizar las evidencias escaladas por L2, aplicar procedimiento de respuesta ante incidentes adecuado. Podríamos estar ante una máquina comprometida, un compromiso de cuenta, pérdida de información, … Esto se determinará dentro del nivel 3. Se aplicarán medidas tanto de contención, como de erradicación y recuperación.	L3

Tabla 95. Procedimiento UC04-MAIL-001

Notas importantes a tener en cuenta:

- Hay que tener en cuenta que estos procedimientos son aproximativos y por tanto evolutivos.
- La parte más compleja de documentar y procedimentar es la parte del L2. De hecho, dado que es un trabajo de investigación, se pueden dar ciertas pautas, pero no indicar qué hacer en cada paso. No son procesos tan automatizados.

- Las acciones descritas como de bloqueo pueden llegar a automatizarse a través del uso de APIs y mejor aún, mediante la implementación de soluciones SOAR.
- Dependiendo del expertise del equipo, habrá funciones del nivel 2 que puedan llegar a ser realizadas por el nivel 1. Esto generará un beneficio al SOC.
- La acción del nivel 3 es la definida dentro de los equipos DFIR. Dichos procesos deben ser aprobados a nivel organizativo.
- Es conveniente que se separen roles como son los de los analistas, con la gente que se encarga de administrar la solución SIEM y la parte de inteligencia.
- Los tiempos de respuesta SLA se aplicarán al nivel 1 especialmente, incrementándose para el nivel 2, sin existir para el nivel 3.

4. Conclusiones

Conclusiones del trabajo.

A lo largo del presente TFM se ha logrado demostrar las ventajas que supone el tener implementado un sistema SIEM dentro de una organización. Además, se ha podido ver hacia donde van encaminados los esfuerzos de las empresas que desarrollan este tipo de sistemas, y qué ventajas nos van a proporcionar.

Se ha podido ver también, que es factible por un coste reducido, el tener una solución SIEM con cierta madurez. Sí hay que indicar, que dada la madurez de estos sistemas, es imposible el crear una solución personalizada que se acerque en sus capacidades a las soluciones comerciales, con un coste en recursos humanos y monetarios que compensen el no contratar un servicio profesional.

También ha quedado demostrado que, el SIEM tiene que ser una herramienta que nos facilite nuestra labor dentro del campo de la seguridad, pero de momento no pueden sustituir la creación de la Inteligencia a nivel de seguridad corporativa. Gracias al machine learning, se van a conseguir detectar más fácilmente ciertos tipos de amenazas, al igual que gracias a los SOAR, se va a poder ser más rápidos en las respuestas ante incidentes en sus fases iniciales. Pero ninguna de las dos características hace que tengamos que excluir la importancia de las políticas de seguridad, del desarrollo de una arquitectura robusta, o del expertise de los analistas a la hora de detectar lo que cada día es más habitual, que son ataques más complejos y evolucionados.

Logros obtenidos.

Al final hemos podido lograr el objetivo principal que se buscaba con el presente trabajo. Por un lado, se ha conseguido transmitir la situación actual y futuro cercano de las soluciones SIEM. Por otro, se ha mostrado detalladamente la forma en que una organización puede aunar un Data Lake y un SIEM dentro de una organización a coste relativamente bajo.

Si es cierto que, este ahorro en costes siempre tiene un lado negativo. Y este es que el trabajo de creación e implementación se multiplica, sin llegar a cotas de funcionalidad que nos da una solución comercial de última generación.

Esto último ha hecho que obviamente no se haya podido implementar un SIEM enteramente funcional. Pero esto en sí es obvio, dado que ya solo la implementación de la arquitectura del Elastic como de Splunk, llevaría mucho tiempo y recursos dedicados dentro de una organización. No hay que olvidar que, existen empresas en parte dedicadas a realizar estas funciones, y que se tardan muchas jornadas en poder implementarlo dentro de una organización. A esto hay que sumarle las horas de creación de la inteligencia a añadir al SIEM. La creación de los eventos notables, alertas, Dashboards, Playbooks, … deben de ser desarrollados por un equipo especializado y multidisciplinar dentro del

campo de la TI. Esto obviamente no puede quedar reflejado dentro de un TFM limitado a tan pocas horas.

A modo de resumen, por tanto se puede llegar a la conclusión de que se ha cumplido completamente con los objetivos iniciales que se buscaban con la realización del proyecto.

Planificación y seguimiento del trabajo.

A nivel de planificación, se ha conseguido respetar los tiempos que se habían marcado inicialmente. El tener claro lo que había que hacer, y la forma de hacerlo, ha ayudado en su consecución.

Sí es cierto que como se ha comentado antes, se cumplen correctamente con los tiempos si consideramos que el objetivo del proyecto es dejar constancia de una metodología un marco de trabajo futuro. Es obvio que, si se hubiera pretendido realizar un proyecto completo de una solución SIEM para ser puesta en producción, no hubiera sido ni realista ni factible el haberlo planteado dentro de un TFM.

Líneas de trabajo futuro.

Acabamos de ver que, la implementación de un sistema SIEM eficiente, es muy complejo.

Lo más importante a la hora de desarrollar un sistema SIEM es la aplicación de inteligencia. Inteligencia a nivel de determinar qué fuentes y con qué información deben de ser indexadas dentro del sistema. Inteligencia a la hora de determinar de entre todos los eventos, cuáles son realmente interesantes para ser considerados como eventos notables. E inteligencia a la hora de determinar qué alertas se pueden generar en base a estos últimos.

Todo el sistema es como un ente vivo, en el que la mejora continua y la evaluación debe de ser un objetivo constante a conseguir. Tanto los sistemas, como las amenazas evolucionan a lo largo del tiempo, y debemos adaptarnos a dichos cambios, intentando incluso anticiparnos en ciertos aspectos.

Sí es cierto, que la complejidad de los sistemas SIEM están alcanzado cotas tan altas, que sin apoyarnos en soluciones comerciales, va a ser muy difícil el poseer soluciones eficaces.

Esto último se ve perfectamente en el uso del machine learning, y de los sistemas SOAR. Sobre todo el primero, no es factible el poder adaptarlo a un solución open source debido a su complejidad, y a que en el fondo, el SIEM es una herramienta más del SOC, que nos facilita el objetivo último de securizar y proteger una organización. Pasa exactamente lo mismo a nivel de automatización de respuesta, si bien los SOAR se basan en programación a nivel del API, el trabajo que debería de desarrollarse a nivel de desarrollo e implementación dentro de una organización para crear algo parecido a esto

sería tan grande que debería de haber un único equipo dedicado en exclusividad. Esto no tiene sentido a nivel de costes-beneficio.

Por tanto, el planteamiento futuro del presente trabajo debe de estar destinado a fomentar y desarrollar los aspectos relacionados con la inteligencia, adaptándolos a aquellos lugares donde normalmente un SIEM genérico no llega. Es decir, adaptándolo a las propias necesidades de la organización donde se está implementando esta solución.

En base a todo esto, se deben abordar como acción primaria, la determinación de todos los eventos notables junto con sus alertas, que se pueden implementar en el sistema donde se desarrolle. Buscando en todo momento cubrir todas las posibles debilidades organizativas existentes, y centrándose en experiencias pasadas y posibles objetivos futuros, para proteger de forma proactiva la organización.

Y en base a lo anterior, se deberán crear nuevas alertas dentro de Splunk, desarrollando sus correspondientes procedimientos de actuación.

También se deberán desarrollar y profundizar todo lo relacionado con el enriquecimiento de las alertas a través de los knowledge objects de Splunk, y las posibilidades de automatización que nos dan el uso de ciertas Apps.

Esta implementación, junto con el ahorro en costes del planteamiento, más un equipo altamente cualificado, hará que la organización que implemente esta solución se acerque a niveles de madurez elevados, sin un gran coste económico. Siempre entendiendo, que será muy complicado llegar al nivel que se tendría con una solución SIEM comercial de tercera generación.

5. Glosario

Apache Cassandra. Es una base de datos NoSQL distribuida y basada en un modelo de almacenamiento de tipo clave-valor. Está programado en Java, siendo de código abierto. Permite grandes volúmenes de datos en forma distribuida. El objetivo es la escalabilidad lineal y la disponibilidad. Los nodos se comunican entre ellos a través de un protocolo P2P, esto hace que la redundancia sea máxima.
Apache Hadoop. La biblioteca de software Apache Hadoop es un framework que permite el procesamiento distribuido de grandes conjuntos de datos dentro de clústeres de computadores utilizando modelos de programación simples.
API. Acrónimo del inglés application programming interface, es un conjunto de subrutinas, funciones y procedimientos que ofrece cierta biblioteca para ser usado por otro software como una capa de abstracción. Otra definición es la de un conjunto de definiciones y protocolos que se utiliza para desarrollar e integrar el software de las aplicaciones. Las API permiten que los programas y servicios se comuniquen con otros, sin necesidad de saber cómo están implementados.
APT. Malware creado específicamente para atacar a una empresa o gobierno concretos con el objetivo principal de robar información (aunque puede tener otros) y permanecer ocultos el mayor tiempo posible.
BI. Se define Businness Intelligence al conjunto de estrategias, aplicaciones, datos, productos, tecnologías y arquitecturas técnicas, los cuales están enfocados a la administración y creación de conocimiento sobre el medio, esto se consigue mediante el análisis de todos los datos existentes dentro de una organización.
Big Data. Se trata de un término común en el que se agrupan todas las técnicas de tratamiento de grandes volúmenes de datos, que no se realizan mediante las metodologías de análisis y herramientas clásicas. Gartner lo define como datos que contienen mayor variedad y que se presentan en volúmenes crecientes y a una velocidad superior. Se introduce por tanto eo concepto de las tres V: Volumen, velocidad y variedad.
CISO. Del inglés Chief Information Security Officer, es el director de seguridad de la información. Su rol se basa en un desempeño a nivel ejecutivo y su función principal es la de alinear la seguridad de la información con los objetivos de negocio. Se garantiza en todo momento que la información de la empresa está protegida adecuadamente.
Cyber Kill Chain. Concepto acuñado por Lockheed Martin Corporation, es un término basado en tácticas militares que ayuda a entender cómo operan los atacantes. Se define como una cadena dado que, para la obtención del objetivo por parte del atacante, debe de dar una serie de pasos en un orden determinado. Hay un total de siete pasos: reconocimiento, preparación, distribución, explotación, instalación, comando y control y acciones sobre los objetivos.
Data Center. Es el espacio donde se concentran los recursos necesarios para el procesamiento de la información de una organización. Este espacio puede ser físico o virtual. Están compuestos por los siguientes elementos y funcionamientos: servidores, conectividad de red, energía, climatización, monitorización y sistemas de seguridad.

Data Lake. Es un repositorio centralizado que permite almacenar todos los datos estructurados o no a cualquier escala. Estos datos se pueden almacenar sin necesidad de estructurarlos para posteriormente procesarlos, visualizarlos, análisis de datos en tiempo real y aprendizaje automático entre otros. El Data Lake se basa en Big Data y su mayor virtud es su potencialidad en tiempo real.
DLP. Es un software que detecta potenciales brechas o exfiltraciones de datos, o realiza funciones de prevención mediante la monitorización, detección y el bloqueo de datos considerados sensibles tanto en los sistemas finales de usuario, como a lo largo de la red interna o en los almacenes de datos.
EDR. Acrónimo del inglés de Endpoint Detection and Response. Este tipo de dispositivos complementan a los ya existentes dentro de las organizaciones como los EPP (Endpoint Protection Plataform), llegando a cubrir necesidades que estos últimos no consiguen llegar, robusteciendo la seguridad corporativa. Estos sistemas se basan en detener, responder y remediar. Combaten las amenazas avanzadas y responden ante los incidentes de seguridad en los clientes finales. Poseen cualidades como el análisis de comportamiento, control de aplicaciones, monitorización de la red y la respuesta ante los incidentes. También puede llegar a proporcionar detalles forenses.
ElasticSearch. Se trata de un servidor de búsqueda basado en Lucene. Posee un motor de búsqueda de texto complejo, distribuido y con capacidad de multitenencia con un interfaz web RESTfull y basado en documentos JSON. Está desarrollado en Java y está publicado como código abierto.
Gartner. Es una empresa consultora y de investigación sobre las tecnologías de la información americana. Realiza funciones de investigación tecnológica, análisis y consultoría.
IaaS. Acrónimo de Infraestructure as a Service, se trata de una infraestructura totalmente gestionada, alojada en la nube y disponible a las necesidades de los clientes. Se adapta a nivel de arquitectura y recursos a la demanda del cliente en cada momento. Y en base a lo consumido, es lo que se paga. Permite un mayor control que otros tipos de soluciones en la nube como puede ser el modelo PaaS, de este modo es el cliente el responsable de todo lo relacionado con el mantenimiento de la infraestructura, incluyendo el escalado a nivel de aplicación en función de sus necesidades. Por tanto, esta solución se basa en que un proveedor proporciona a los clientes acceso de pago por uso del almacenamiento, uso de redes, de los servidores y otros recursos informáticos alojados en la nube.
IDS. Se trata de un sistema de detección de intrusos (IDS del inglés, Intrusion Detection System) es un programa de detección de accesos no autorizados a un sistema o red. Los IDS poseen sensores o sondas de los que se obtiene los datos externos, dichos sensores esnifan el tráfico y el núcleo del IDS es el que realiza el trabajo de detección. Si existe bloqueo ante la detección de una amenaza se denominan IPS (Intrusion Prevention System).
IOA. Acrónimo de Indicator of Attack, se centran en detectar la intención de un atacante sobre un objetivo, independientemente del malware o exploit que sea usado. Son por tanto las serie de acciones que un adversario debe de realizar para tener éxito. Los IOAs no se centran en las herramientas específicas que se utilizan para lograr los objetivos, sino que se centran en los pasos, la intención del adversario y los resultados que está tratando de lograr. Mediante los IOAs se consigue detectar un ataque antes de que los IOCs se hagan

presentes. Se busca de manera proactiva la detección del ataque y su bloqueo antes de que se produzca un impacto significativo.

IOC. Acrónimo de Indicators of Compromise, hace referencia a una tecnología estandarizada que consiste en definir las características técnicas de una determinada amenaza por medio de las evidencias existentes en un equipo comprometido, de manera que puedan servir para identificar otros sistemas afectados por la misma amenaza o prevenirse de la misma.

IoT. Se trata del Internet de las cosas. Es un concepto que se refiere a la interconexión digital de objetos cotidianos conectados a Internet o a una red privada. Entra en juego el concepto de interacción M2M (machine to machine), que es la capacidad de estos dispositivos de interactuar entre ellos sin interacción humana.

MISP. Proyecto open source que proporciona una plataforma de inteligencia de amenazas. Se encarga básicamente de la compartición de indicadores de compromiso con la comunidad en la que se puede colaborar.

MongoDB. Es una base de datos distribuida, basada en documentos y de uso general que se ha diseñado para desarrollar aplicaciones modernas y para aplicaciones orientadas en la nube. Almacena los datos en forma de documentos tipo JSON. Además, posee un potente lenguaje de búsqueda. Este esquema de almacenamiento es dinámico, haciendo que la integración de los datos en ciertas aplicaciones sea más fácil y rápida.

MSSP. Acrónimo de Managed Security Service Provider, son proveedores de servicios de seguridad gestionada. Son capaces de ofrecer una visión integral y multidisciplinar de la seguridad corporativa en cada una de las áreas de la organización. Proporcionan una seguridad preventiva y proactiva del ecosistema de la empresa. Para ello gestionan la seguridad de las infraestructuras de la organización, sirviéndose de herramientas especializadas capaces de detectar brechas de seguridad, siendo capaces a su vez de simular escenarios reales de ataques anticipándose a las amenazas. Sus funciones son por tanto prevenir, detectar, responder y predecir.

NoSQL. Se trata de una amplia clase de sistemas de gestión de base de datos que difieren del modelo clásico de SGBDR en aspectos muy importantes, el más destacado es que no usan sentencias SQL como lenguaje principal de consultas (pero sí pueden sopórtalo). Los datos almacenados no requieren estructuras fijas como son las tablas, normalmente no soportan operaciones JOIN, ni garantizan completamente el ACID (atomicidad, consistencia, aislamiento y durabilidad), y habitualmente escalan bien a nivel horizontal.

UEBA. Acrónimo de User and Entity Behaviour Analytics, es el aprendizaje automático basado en IA para identificar cambios en los patrones de comportamiento de los usuarios o entidades dentro de una organización, y que pueden ser identificados como un posible riesgo. Estos dispositivos aprenden del comportamiento de los agentes estableciéndose una línea base y aplicando porcentajes de riesgo adaptativo a lo largo del tiempo, detectar posibles incidentes de seguridad.

SaaS. Acrónimo de Software as a Service, es un modelo de distribución de software donde el soporte lógico de éste y los datos que maneja son alojados en servidores externos a la organización que lo contrata, al que se accede vía Internet. La empresa prestadora del servicio se ocupa del mantenimiento, de la operativa diaria y del soporte del software usado por el cliente.

SANS Institute. Es una institución con ánimo de lucro americana. Sus principales objetivos son reunir información sobre todo lo referente a la seguridad informática, y ofrecer capacitaciones y certificaciones en el ámbito de la ciberseguridad.
SEM. Acrónimo de Security Event Management, es un sistema que se encarga del monitoreo en tiempo real, correlación de eventos, notificaciones y vistas de la consola de gestión de la información de seguridad.
SIEM. Acrónimo de Security Information and Event Management, es un sistema centralizado de almacenamiento, junto con la interpretación de lo datos relevantes para la seguridad. Permite un análisis referente a la seguridad de la organización desde un lugar centralizado, facilitando por ello la detección de tendencias y patrones no habituales.
SIM. Acrónimo de Security Information Management, es un sistema encargado del almacenamiento a largo plazo, el análisis y la comunicación de los datos de seguridad.
SOC. Se trata del centro de operaciones de seguridad (Security Operation Center), que previene, monitoriza y controla la seguridad de los sistemas monitorizados.
Threat actor. También conocido como un actor malicioso es una persona o entidad que es responsable de un evento o incidente que tiene la posibilidad de impactar o impacta sobre la seguridad de otra entidad.
TTP. Definido por Mitre, son las tácticas, técnicas y procedimientos desarrollados por los adversarios. Es una forma moderna de ver los ciberataques. En lugar de centrarse en las consecuencias de un ataque a través de los IOCs, se recomienda que los analistas entiendan las tácticas, técnicas y procedimientos de los atacantes.

6. Bibliografía

Fuentes consultadas capítulo 2:

https://es.wikipedia.org/wiki/Gestión_de_información_y_eventos_de_seguridad
https://sofecom.com/que-es-un-siem/
https://en.wikipedia.org/wiki/Security_event_manager
https://en.wikipedia.org/wiki/Security_information_and_event_management
https://en.wikipedia.org/wiki/Security_information_management
https://www.eventtracker.com/EventTracker/media/EventTracker/Files/Solution-Briefs/SolBrief-SANS-Top-20-CIS.pdf
https://www.cisecurity.org/controls/
https://www.rapid7.com/solutions/compliance/critical-controls/
https://www.sans.org/reading-room/whitepapers/analyst/implementing-20-critical-controls-security-information-event-management-siem-systems-34965
https://www.gmdit.com/NewsView.aspx?ID=9IfB2Axzeew=
https://www.splunk.com/pdfs/ebooks/7-siem-trends-to-watch-in-2019.pdf
https://searchdatacenter.techtarget.com/es/definicion/Gestion-de-eventos-e-informacion-de-seguridad-SIEM
https://acis.org.co/portal/sites/all/themes/argo/assets/img/Pagina/PresentacionArmandoCarvajal-LaevoluciondelossistemasSIEMdeseguridaddelainformacion.pdf
https://www.elevenpaths.com/es/noticias-y-eventos/elevenpaths-talks/los-siem-y-la-correlacion-de-eventos-avanzada/index.html#
https://resources.infosecinstitute.com/what-is-a-siem/#gref
https://community.microfocus.com/t5/Security-Blog/SIM-SEM-and-SIEM-Definitions-and-Choosing-the-Right-Enterprise/ba-p/1794733
https://en.wikipedia.org/wiki/Managed_security_service
https://blog.es.logicalis.com/seguridad/soluciones-siem-de-seguridad-las-4-ventajas-mas-importantes
https://www.its-security.es/siem-security-information-event-management-que-es-en-realidad/
https://manuelvieda.com/blog/administracion-de-logs/
https://es.wikipedia.org/wiki/Valores_separados_por_comas
https://es.wikipedia.org/wiki/JSON
https://www.bitlyft.com/understanding-the-difference-of-cloud-vs-on-premise-siem-services/
https://sommet.mx/blog/que-es-user-and-entity-behavior-analytics-ueba
https://es.logrhythm.com/products/
https://docs.splunk.com/Splexicon
https://www.rsa.com/en-us/products/threat-detection-response
https://www.techopedia.com/definition/31778/contextual-data
https://en.wikipedia.org/wiki/Out_of_the_box_(feature)
https://www.gartner.com/reviews/market/security-information-event-management
https://www.linkedin.com/pulse/implantaci%C3%B3n-y-despliegue-siem-security-information-txus
https://es.slideshare.net/pathinishanth/siem-architecture
https://en.wikipedia.org/wiki/Log_file

https://www.loggly.com/docs/log-retention/
https://en.wikipedia.org/wiki/User_behavior_analytics
https://resources.infosecinstitute.com/what-is-a-siem/#gref
https://www.gartner.com/en/documents/3834694
https://www.interempresas.net/Electronica/Articulos/232650-Integracion-SIEM-SOC-Ciberseguridad-privacidad-motores-clave-esencia-accesibilidad.html
https://blog.logsign.com/security-information-and-event-management-architecture/
https://www.youtube.com/watch?v=mGfMrXW8W98
https://www.exabeam.com/siem-guide/
https://www.exabeam.com/siem/2019-gartner-peer-insights-voice-of-the-customer-siem-exabeam/
https://it.umich.edu/information-technology-policies/general-policies/DS-19
https://www.youtube.com/watch?v=FUgScyM9bkM
https://resources.infosecinstitute.com/category/enterprise/threat-hunting/threat-hunting-benefits/threat-hunting-vs-siem/#gref
https://www.pandasecurity.com/spain/mediacenter/adaptive-defense/threat-hunting-por-que-necesario/
https://ibm-security-solutions-master-threat-hunting-ebook.mybluemix.net/
https://www.exabeam.com/product/exabeam-incident-responder/
https://www.exabeam.com/product/exabeam-threat-hunter/
https://www.gartner.com/reviews/market/security-information-event-management
https://www.youtube.com/watch?v=HDE9MF2MiHc
https://www.gartner.com/en/information-technology/glossary/security-information-and-event-management-siem

Fuentes consultadas capítulo 3:

https://www.splunk.com/en_us/resources/search-processing-language.html
https://docs.splunk.com/Splexicon:SPL
https://www.yelp.com/engineering
https://elastalert.readthedocs.io/en/latest/
https://github.com/Yelp/elastalert
https://www.theninjacto.xyz/ElastAlert-consultas-Elasticsearch-Alertas-Controller/
https://www.elastic.co/guide/en/elasticsearch/reference/7.4/query-dsl-range-query.html
https://www.elastic.co/guide/en/elasticsearch/reference/7.4/query-dsl-term-query.html
https://www.elastic.co/guide/en/elasticsearch/reference/7.4/query-dsl-match-query.html
https://www.elastic.co/guide/en/elasticsearch/reference/7.4/query-dsl.html
https://www.elastic.co/guide/en/elasticsearch/reference/7.4/query-dsl-bool-query.html
https://www.elastic.co/guide/en/elasticsearch/reference/7.4/query-filter-context.html
https://www.elastic.co/es/products/beats/filebeat
https://www.elastic.co/es/products/beats/metricbeat
https://www.elastic.co/es/products/beats/packetbeat

https://www.elastic.co/es/products/beats/winlogbeat
https://www.elastic.co/es/products/beats/auditbeat
https://www.elastic.co/es/products/beats/heartbeat
https://www.elastic.co/es/products/beats/functionbeat
https://www.elastic.co/es/products/beats
https://www.elastic.co/es/products/kibana/features
https://www.elastic.co/es/products/kibana
https://www.adictosaltrabajo.com/2015/12/27/introduccion-a-kibana/
https://openwebinars.net/blog/que-es-elk-elasticsearch-logstash-y-kibana/
https://www.elastic.co/es/what-is/elk-stack
https://www.elastic.co/es/products/logstash
https://www.ochobitshacenunbyte.com/2018/08/28/que-es-y-como-funciona-elasticsearch/
https://docs.splunk.com/Documentation/Forwarder/7.3.2/Forwarder/Abouttheuniversalforwarder
https://docs.splunk.com/Documentation/Splunk/8.0.0/Data/Extractfieldsfromfileswithstructureddata
https://docs.splunk.com/Splexicon:Universalforwarder
https://docs.splunk.com/Splexicon:Forwarder
https://docs.splunk.com/Splexicon:Heavyforwarder
https://www.splunk.com/blog/2016/12/12/universal-or-heavy-that-is-the-question.html
https://www.edureka.co/blog/splunk-architecture/
https://docs.splunk.com/Splexicon:Indexer
https://docs.splunk.com/Splexicon:Distributedsearch
https://docs.splunk.com/Splexicon:Indexercluster
https://docs.splunk.com/Splexicon:Searchpeer
https://www.learnsplunk.com/splunk-architecture.html
https://docs.splunk.com/Splexicon:Searchhead
https://docs.splunk.com/Splexicon:Licensemaster
https://docs.splunk.com/Documentation/Splunk/7.3.2/Updating/Forwardermanagementoverview
https://docs.splunk.com/Splexicon:Licenseslave
https://docs.splunk.com/Splexicon:Masternode
https://docs.splunk.com/Splexicon:Peernode
https://docs.splunk.com/Documentation/Splunk/6.4.3/Indexer/Basicclusterarchitecture
https://docs.splunk.com/Documentation/Splunk/7.3.2/SearchReference/UnderstandingSPLsyntax
https://www.aditumpartners.com/splunk-common-information-model/
https://docs.splunk.com/Documentation/CIM/4.13.0/User/Overview
https://github.com/hire-vladimir/SA-cim_vladiator
https://docs.splunk.com/Documentation/PCI/3.8.1/Install/Notableevents
https://www.youtube.com/watch?v=ItzjnkVxBeE
https://www.youtube.com/watch?v=luLb1Y0gsSg
https://www.youtube.com/watch?v=xtyH_6iMxwA
https://www.youtube.com/watch?v=hffwH4XT0z8
https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781788835237/1/ch01lvl1sec17/receiving-data-using-the-http-event-collector
https://www.youtube.com/watch?v=qROXrFGqWAU

https://www.youtube.com/watch?v=9awwyjORWO8
https://en.wikipedia.org/wiki/Kill_chain
https://www.csacademy.nl/images/scripties/2018/Paul-Pols---The-Unified-Kill-Chain.pdf
https://nomagic.software/post/2018-11-10-the-kill-chain/
https://www.linkedin.com/pulse/estrategias-y-t%C3%A1cticas-de-defensa-frente-ciberataques-alejandro/
https://www.welivesecurity.com/la-es/2019/06/06/como-utilizar-mitre-attck-repositorio-tecnicas-procedimientos-ataques-defensas/
https://mitigatehub.com/the-unified-kill-chain-part-2/
https://sgros-students.blogspot.com/2019/01/mitre-att-and-unified-kill-chain.html
https://medium.com/mitre-attack/att-ck-101-17074d3bc62
https://github.com/Neo23x0/sigma
https://www.slideshare.net/secret/gvgxeXoKblXRcA
https://ciberseguridad.blog/que-son-las-reglas-sigma-y-por-que-las-necesitas/
https://n0where.net/generic-signature-format-for-siem-systems-sigma
https://socprime.com/en/blog/sigma-rules-guide-for-arcsight/
http://www.redseguridad.com/especialidades-tic/otros/adoptando-el-att-ck-de-mitre-para-threat-hunting
https://www.atlassian.com/incident-management/kpis/common-metrics
https://searchdatacenter.techtarget.com/es/cronica/Desafios-y-beneficios-de-usar-el-marco-de-Mitre-ATTCK
https://www.anomali.com/es/what-mitre-attck-is-and-how-it-is-useful
https://docs.splunk.com/Documentation/SplunkCloud/8.0.0/Data/UsetheHTTPEventCollector
https://wiki.splunk.com/images/e/e0/Splunk-Icon-Styleguide.pdf
https://wiki.splunk.com/Main_Page
https://answers.splunk.com/storage/temp/174425-splunk-common-ports.png
https://www.splunk.com/en_us/data-insider/what-is-siem.html
http://www.reydes.com/d/?q=Categorias_de_Eventos_de_Seguridad_en_Windows
https://docs.microsoft.com/en-us/windows-server/identity/ad-ds/plan/appendix-l--events-to-monitor
https://docs.microsoft.com/en-us/windows/security/threat-protection/use-windows-event-forwarding-to-assist-in-intrusion-detection
https://www.xplg.com/windows-server-security-events-list/
https://www.petri.com/monitoring-windows-event-logs-for-security-breaches
https://docs.microsoft.com/es-es/dotnet/framework/wcf/feature-details/auditing-security-events
https://docs.microsoft.com/en-us/windows-server/identity/ad-ds/plan/appendix-l--events-to-monitor
https://www.beyondtrust.com/blog/entry/windows-server-events-monitor
https://www.welivesecurity.com/la-es/2016/09/08/eventos-de-windows-analisis-forense/
https://support.microsoft.com/es-es/help/977519/description-of-security-events-in-windows-7-and-in-windows-server-2008
https://www.dnsstuff.com/windows-event-log-monitoring-best-practices
https://www.xplg.com/windows-server-security-events-list/
https://www.loggly.com/ultimate-guide/windows-logging-basics/

https://www.howtogeek.com/school/using-windows-admin-tools-like-a-pro/lesson3/
https://support.microsoft.com/en-us/help/323076/how-to-set-event-log-security-locally-or-by-using-group-policy
https://hackernoon.com/the-windows-event-forwarding-survival-guide-2010db7a68c4
https://docs.microsoft.com/en-us/windows/win32/winrm/portal
https://www.arsys.es/blog/soluciones/edr-seguridad/
https://www.magazcitum.com.mx/?p=3739#.XdGC61dKjIU
https://fwhibbit.es/sysmon-el-gran-hermano-de-windows-y-el-super-sysmonview
https://answers.splunk.com/answers/499999/how-to-setup-a-channel-to-send-raw-data-to-http-ev.html
https://docs.splunk.com/Documentation/Splunk/latest/Data/UsetheHTTPEventCollector
https://docs.splunk.com/Documentation/SplunkCloud/8.0.0/Data/AboutHECIDXAck
https://manualkibanaocds.readthedocs.io/es/latest/C2/Seccion1/3_Kibana.html
https://manualkibanaocds.readthedocs.io/es/latest/C2/Seccion1/1_ElasticSearch.html
https://manualkibanaocds.readthedocs.io/es/latest/C2/Seccion1/2_Logstash.html
https://www.adictosaltrabajo.com/2016/06/01/administracion-de-elasticsearch/
http://www.arquitectoit.com/elasticsearch/conceptos-basicos-elasticsearch/
https://www.elastic.co/guide/en/elasticsearch/reference/current/modules-node.html#data-node
https://www.elastic.co/es/blog/implementing-hot-warm-cold-in-elasticsearch-with-index-lifecycle-management
https://manualkibanaocds.readthedocs.io/es/latest/C2/Seccion1.html

Fuentes imágenes:

Ilustración 3: Cuadrante mágico de Gartner octubre 2018. Fuente: Rapid7, https://www.rapid7.com/contentassets/103ddfbbf31943f183d6344f33a9c2b0/2018-siem-mq.jpg
Ilustración 4: Diferencias según Gartner. Fuente: Pandasecurity, https://www.pandasecurity.com/spain/mediacenter/src/uploads/2018/11/Threat-Hunting-vs-detection.jpg
Ilustración 5. Elastic Stack
https://www.elastic.co/static-res/images/elk/elk-stack-elkb-diagram.svg
Ilustración 6. Fases CKC
https://img.deusm.com/darkreading/MarilynCohodas/killchainchart.jpg
Ilustración 7. Unified Kill Chain
https://upload.wikimedia.org/wikipedia/commons/c/c2/The_Unified_Kill_Chain.png
Ilustración 8. Matriz ATT&CK
https://miro.medium.com/max/1200/0*d7eYAV-RxXPYXB7E.jpg
Ilustración 9. Sigma Rules
https://github.com/Neo23x0/sigma/raw/master/images/Sigma-description.png

Fuentes Glosario:

https://hadoop.apache.org/
https://es.wikipedia.org/wiki/Apache_Hadoop
https://aws.amazon.com/es/big-data/datalakes-and-analytics/what-is-a-data-lake/
https://artyco.com/data-warehouse-data-lake-que-es/
https://en.wikipedia.org/wiki/Threat_actor
https://en.wikipedia.org/wiki/Malware_Information_Sharing_Platform
https://es.wikipedia.org/wiki/Software_como_servicio
https://es.wikipedia.org/wiki/Software_de_prevenci%C3%B3n_de_p%C3%A9rdida_de_datos
https://en.wikipedia.org/wiki/Data_loss_prevention_software
https://es.wikipedia.org/wiki/Amenaza_persistente_avanzada
https://www.ibm.com/blogs/think/es-es/2015/02/24/cuidado-con-las-apts/
https://latam.kaspersky.com/blog/que-es-apt/761/
https://www.kaspersky.es/blog/que-es-una-apt/966/
https://es.wikipedia.org/wiki/Inteligencia_empresarial
https://www.incibe-cert.es/blog/indicadores-de-compromiso
https://www.crowdstrike.com/blog/indicators-attack-vs-indicators-compromise/
https://qbhackers.com/soc-indicator/
https://en.wikipedia.org/wiki/Data_loss_prevention_software
https://www.vernegroup.com/vernetelecom/actualidad/conoces-la-figura-del-mssp-managed-security-service-provider-en-las-empresas
https://en.wikipedia.org/wiki/Managed_security_service
https://es.wikipedia.org/wiki/Software_como_servicio
https://azure.microsoft.com/es-es/overview/what-is-saas/
https://www.oracle.com/es/big-data/guide/what-is-big-data.html
http://www.iic.uam.es/big-data/
https://es.wikipedia.org/wiki/Macrodatos
https://www.masterbigdataucm.com/que-es-big-data/
https://es.wikipedia.org/wiki/MongoDB
https://www.mongodb.com/es
https://www.welivesecurity.com/la-es/2014/08/07/enfoque-militar-gestion-seguridad/
https://www.incibe-cert.es/blog/cyber-kill-chain-sistemas-control-industrial
https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf
https://www.pandasecurity.com/rfiles/enterprise/solutions/ad360/1704-WHITEPAPER-CKC-ES.pdf
https://www.datos101.com/blog/que-es-un-datacenter/
https://es.wikipedia.org/wiki/Centro_de_procesamiento_de_datos
https://www.arubanetworks.com/es/productos/seguridad/ueba/
https://ciberseguridad.blog/ueba-user-and-entity-behavior-analytics-deteccion-por-comportamiento/
https://es.wikipedia.org/wiki/Apache_Cassandra
https://es.wikipedia.org/wiki/Elasticsearch
https://car.mitre.org/Glossary.html
https://www.ibm.com/es-es/cloud/learn/iaas-paas-saas

https://www.axarnet.es/blog/saas-paas-iaas/
https://azure.microsoft.com/es-es/overview/what-is-iaas/
https://www.exabeam.com/information-security/what-is-mitre-attck-an-explainer/
https://es.wikipedia.org/wiki/Gestión_de_información_y_eventos_de_seguridad
https://www.redhat.com/es/topics/api/what-are-application-programming-interfaces
https://es.wikipedia.org/wiki/Interfaz_de_programaci%C3%B3n_de_aplicaciones
https://es.wikipedia.org/wiki/NoSQL
https://www.incibe.es/protege-tu-empresa/blog/ceo-ciso-cio-roles-ciberseguridad
https://es.wikipedia.org/wiki/Gartner_(empresa)
https://es.wikipedia.org/wiki/Centro_de_operaciones_de_seguridad
https://es.wikipedia.org/wiki/Sistema_de_detecci%C3%B3n_de_intrusos
https://www2.deloitte.com/es/es/pages/technology/articles/IoT-internet-of-things.html
https://es.wikipedia.org/wiki/Internet_de_las_cosas
https://www.kaspersky.es/blog/epp-edr-importance/16154/
https://www.pandasecurity.com/spain/mediacenter/seguridad/soluciones-edr-tendencia/
https://www.arsys.es/blog/soluciones/edr-seguridad/
http://one-esecurity.com/es/SANS/
https://es.wikipedia.org/wiki/SANS_Institute