Universitat Oberta de Catalunya
uoc.edu

Ventajas e Implementación de un sistema SIEM

Jairo García Merino
Máster Universitario en Seguridad de las Tecnologías de la Información y de las Comunicaciones
Ad-hoc INCIBE

Marco Antonio Lozano
Víctor Garcia Font

4 de junio de 2018

FICHA DEL TRABAJO FINAL

Título del trabajo:	*Ventajas e Implementación de un sistema SIEM*
Nombre del autor:	*Jairo García Merino*
Nombre del consultor/a:	*Marco Antonio Lozano*
Nombre del PRA:	*Víctor Garcia Font*
Fecha de entrega (mm/aaaa):	*06/2018*
Titulación:	*Máster Universitario en Seguridad de las Tecnologías de la Información y de las Comunicaciones*
Área del Trabajo Final:	*M1.830 - TFM-Ad hoc aula 1*
Idioma del trabajo:	*Español*
Palabras clave	*Máximo 3 palabras clave, validadas por el director del trabajo (dadas por los estudiantes o en base a listados, tesauros, etc.)*

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

El acrónimo inglés de SIEM se corresponde con Security Information and Event Management, un sistema de gestión de eventos y seguridad de la información, proporciona un punto común para recoger información. Se aplica a múltiples sistemas y aplicativos, que engloban diferentes tecnologías y fabricantes, teniendo en común que son utilizados dentro de una organización o empresa. La recogida de la información en un punto común en un sistema SIEM, tiene el valor añadido que permite su categorización, de tal forma que se pueda producir un escalado de aquellos eventos más relevantes que requieran de intervenciones adicionales. La eficacia del sistema será mayor cuanto mejores sean los controles que estén definidos, ya que la seguridad de los distintos sistemas estará “vigilada” por estos.

Se han evaluado de forma teórica los conceptos que subyacen detrás de la recopilación de información a través de logs de registro, y se ha desarrollado un SIEM sobre la pila ELK.

Abstract (in English, 250 words or less):

The English acronym of SIEM corresponds to Security Information and Event Management, an event management and information security system, providing a common point to gather information. It is applied to multiple systems and applications, which encompass different technologies and manufacturers, having in common that they are used within an organization or company. The collection of the information in a common point in a SIEM system has the added value that allows its categorization, in such a way that it can produce an escalation of those most relevant events that require additional interventions. The effectiveness of the system will be greater the better the controls that are defined, since the security of the different systems will be "monitored" by them.

The concepts behind the collection of information through log logs have been theoretically evaluated, and a SIEM on the ELK stack has been developed.

Índice

Lista de figuras

1. Introducción

1.1 Contexto y justificación del Trabajo

TFM en empresa: INCIBE. Tras el incidente del 12 de mayo de 2017 (WannaCry) muchas empresas han decidido implantar un servicio de SOC (Security Operations Center) con el objetivo, entre otras cosas, de adelantarse a la materialización de una amenaza. A través de este trabajo, el alumno se familiarizará con este tipo de herramientas analizando las ventajas que supone disponer de una aplicación de estas características y donde además deberá implementar una solución que permita correlar eventos y enviar alertas.

1.2 Objetivos del Trabajo

Este TFM se divide en dos bloques bien diferenciados, el primero se centra en conceptos teóricos sobre los sistemas SIEM y sus ventajas. El segundo bloque será la puesta en marcha de un sistema SIEM, dando una aplicación práctica a los conceptos desarrollados en el primer bloque.

Objetivos teóricos:

- Definición de SIEM y sus ventajas.
- Breve estado del arte.
- Retos del sistema
- Arquitectura de implementación.

Objetivos prácticos:

- Laboratorio e implementación.

ELK es un conjunto de herramientas open source que se enlazan para crear una administración de logs, permitiendo su monitorización, agrupación y análisis. Estas herramientas son:

ElasticSearch - actúa como base de datos y motor de búsqueda
Logstash - actúa como herramienta de carga y transformación
Kibana - actúa como herramienta de visualización de datos

Pueden utilizarse de forma independiente y tienen múltiples aplicaciones en el mercado, pero su funcionalidad en este TFM ha sido para crear un sistema SIEM.

De forma esquemática podemos interpretar una pila con estas aplicaciones, de tal forma que la parte inferior representa el origen, la capa intermedia representa el almacenamiento y la capa superior sería la visualización.

1.3 Enfoque y método seguido

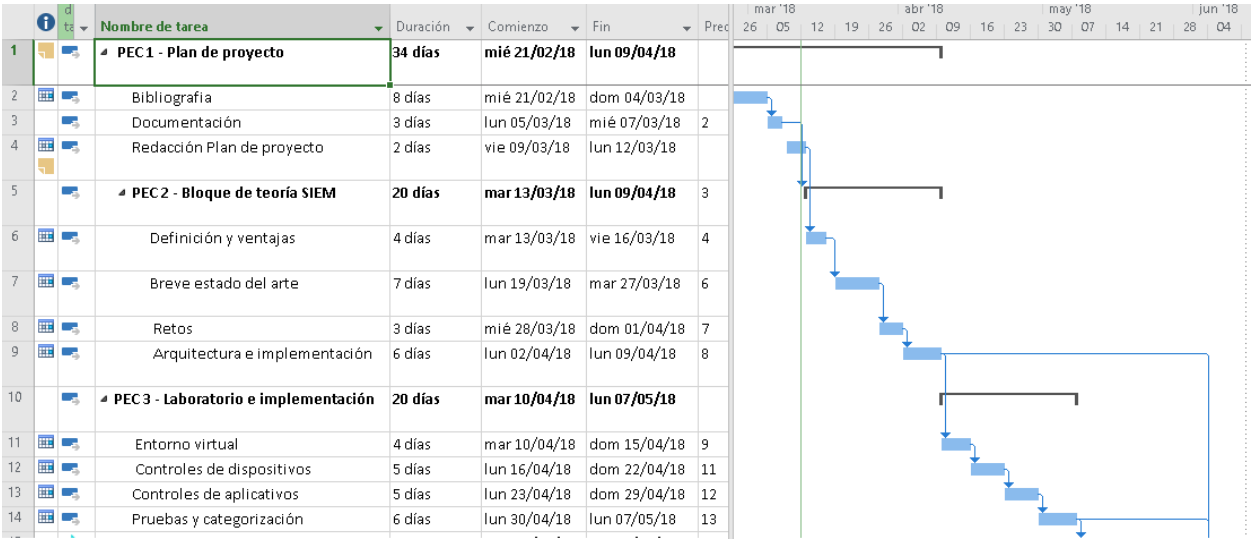
La parte teórica se ha desarrollado estudiando la bibliografía existente sobre la materia, se ha hecho referencia a los principales libros y manuales consultados durante su elaboración.

Para ver las distintas funcionalidades de un sistema SIEM se han preparado un conjunto de módulos, estos serán probados con juegos de datos para estudiar su viabilidad y demostrar el funcionamiento del sistema, tanto de forma individual como en conjunto.

Las pruebas consistirán en una primera parte de configuración, en la que establecerán los parámetros y reglas que se ejecutarán como función del SIEM. Una segunda parte en la que se establecerán los requisitos de los datos a monitorizar, y se cargarán el sistema un juego de datos para realizar la simulación. Finalmente en una tercera parte se comprobará el funcionamiento de la prueba realizada.

1.4 Planificación del Trabajo

Se ha amoldado la planificación temporal teniendo en cuenta las distintas fechas de entrega de las PECs, así como la disponibilidad del alumno. Los objetivos se han programado en el calendario para que puedan alcanzarse en los plazos dados para el TFM. Se adjunta Diagrama de Gantt:



	Nombre de tarea	Duración	Comienzo	Fin	Pred
1	**PEC 1 - Plan de proyecto**	**34 días**	**mié 21/02/18**	**lun 09/04/18**	
2	Bibliografía	8 días	mié 21/02/18	dom 04/03/18	
3	Documentación	3 días	lun 05/03/18	mié 07/03/18	2
4	Redacción Plan de proyecto	2 días	vie 09/03/18	lun 12/03/18	
5	**PEC 2 - Bloque de teoría SIEM**	**20 días**	**mar 13/03/18**	**lun 09/04/18**	3
6	Definición y ventajas	4 días	mar 13/03/18	vie 16/03/18	4
7	Breve estado del arte	7 días	lun 19/03/18	mar 27/03/18	6
8	Retos	3 días	mié 28/03/18	dom 01/04/18	7
9	Arquitectura e implementación	6 días	lun 02/04/18	lun 09/04/18	8
10	**PEC 3 - Laboratorio e implementación**	**20 días**	**mar 10/04/18**	**lun 07/05/18**	
11	Entorno virtual	4 días	mar 10/04/18	dom 15/04/18	9
12	Controles de dispositivos	5 días	lun 16/04/18	dom 22/04/18	11
13	Controles de aplicativos	5 días	lun 23/04/18	dom 29/04/18	12
14	Pruebas y categorización	6 días	lun 30/04/18	lun 07/05/18	13

Las tareas siguen un flujo secuencial, siendo los principales hitos las entregas de las PECs. La elaboración de este Plan de Trabajo se corresponde con el primer hito, por lo que se considera que se ha completado al 100%, el resto de hitos pendientes del TFM son los siguientes:

- PEC 1 - Plan de trabajo, entrega 12/03/2018 (25 horas de trabajo)

- PEC 2 - Bloque de teoria SIEM, entrega 09/04/2018 (100 horas de trabajo)
- PEC 3 - Laboratorio e implementación, entrega 07/05/2018 (100 horas de trabajo)
- PEC 4 - Memoria final, entrega 04/06/2018 (50 horas de trabajo)
- PEC 5 - Video presentación, entrega 11/06/2018 (25 horas de trabajo)

Para indicar mejor la dedicación a parte de la distribución temporal se han hecho asignaciones de horas de trabajo, de tal forma que el TFM se ajuste a la duración en créditos estimada en el plan de estudios, 300 horas de duración total.

El calendario se han tenido en cuenta los días festivos, ya que al compaginar con la vida laboral/personal es imprescindible aprovechar las vacaciones y los fines de semana. También se han dedicado en proporción mayor número de horas a la parte de teoría que a la práctica debido a la experiencia del alumno, ya que a nivel teórico habrá que superar mayores retos que a nivel práctico, dado que los conocimientos previos del alumno son superiores en la implantación de sistemas, que en el de los conceptos teóricos de seguridad que se van a afrontar.

Resumen de las principales tareas:

Nombre	Duración	Comienzo	Fin
PEC 1 - Plan de proyecto	**14 días**	**21 febrero 2018**	**12 marzo 2018**
Bibliografia	8 días	21 febrero 2018	04 marzo 2018
Documentación	3 días	05 marzo 2018	07 marzo 2018
Redacción Plan de proyecto	2 días	09 marzo 2018	12 marzo 2018
PEC 2 - Bloque de teoría SIEM	**20 días**	**13 marzo 2018**	**09 abril 2018**
Definición y ventajas	4 días	13 marzo 2018	16 marzo 2018
Breve estado del arte	7 días	19 marzo 2018	27 marzo 2018
Retos	3 días	28 marzo 2018	01 abril 2018
Arquitectura e implementación	6 días	02 abril 2018	09 abril 2018
PEC 3 - Laboratorio e implementación	**20 días**	**10 abril 2018**	**07 mayo 2018**
Entorno virtual	4 días	10 abril 2018	15 abril 2018
Controles de dispositivos	5 días	16 abril 2018	22 abril 2018
Controles de aplicativos	5 días	23 abril 2018	29 abril 2018
Pruebas y categorización	6 días	30 abril 2018	07 mayo 2018
PEC 4 - Memoria final	**21 días**	**08 mayo 2018**	**04 junio 2018**
Redacción	15 días	08 mayo 2018	27 mayo 2018
Conclusiones	6 días	28 mayo 2018	04 junio 2018
PEC 5 - Video - Presentación	**6 días**	**05 junio 2018**	**11 junio 2018**
Guión	4 días	05 junio 2018	08 junio 2018
Grabación	2 días	08 junio 2018	11 junio 2018

Los costes del proyecto se estiman en 15.875,00 € (quince mil ochocientos setenta y cinco euros).

Detalle:

Elemento	Unidades	Coste
Ordenador personal	1	800,00 €
Horas de trabajo	300	15.000,00 €
Reprografía	3	75,00 €
Software y licencias	5	0,00 €
	Total	15.875,00 €

1.5 Breve sumario de productos obtenidos

Los conceptos teóricos han sintetizado las ventajas teóricas de un sistema SIEM, mientras que el laboratorio muestra como implementar un SIEM con ELK.

Los componentes que se han utilizado son open source, a excepción del Elasticsearch-Xpac, que se ha considerado utilizar durante el tiempo de evaluación de 30 días, ya que permite una rápida de configuración de los roles de usuarios, así como el módulo de alertas Watcher.

Se adjunta una lista de componentes utilizados, junto con los enlaces a la información de la versión:

Elasticsearch 6.2.4 (info)
Kibana 6.2.4 (info)
Logstash 6.2.4 (info)
Beats 6.2.4 (info)
X-Pack for the Elastic Stack [6.2] (info)
Open Java 8 (info)
Python 2.7 (info)

Los componentes principales de ELK utilizan licencia Apache[i], que para su utilización requiere la conservación del aviso de derecho de autor, así como la descarga de responsabilidad, pero no es una licencia copyleft, ya que no requiere la redistribución del código fuente cuando se distribuyen versiones modificadas. Implica que todo el software producido bajo esta licencia, permite al usuario la libertad de usar el software para cualquier propósito, para distribuirlo, modificarlo y distribuir versiones modificadas, bajo los mismos términos de la licencia, sin preocuparse del copiright.

1.6 Breve descripción de los otros capítulos de la memoria

En esta memòria hay dos bloques bién diferenciados, el primero trata los conceptos teóricos de un SIEM y sus ventajas, pasando por una evolución de las herramientas que han generado este nuevo tipo de productos Comerciales. Se aplicant estos conceptos en el caso de una empresa dando las claves para la toma de decisiones para desarrollar, comprar o externalizar este tipo de productos.
La segunta parte trata sobre la Arquitectura ELK y su implementación pràctica como una herramienta SIEM. En los annexos se exponent detalladamente las alertas genreadas como ejercicios para ver la aplicación pràctica del producto.

2. Definición de SIEM y sus ventajas

El acrónimo inglés de SIEM se corresponde con Security Information and Event Management, un sistema de gestión de eventos y seguridad de la información, proporciona un punto común para recoger información.

El término información de seguridad, se aplica a todos aquellos datos susceptibles de ser tratados tanto de forma individual o correlados con otros, que proporcionen un conocimiento sobre la integridad de un sistema. Los datos pueden proveer de multitud de fuentes, tanto programas software como elementos hardware que forman parte de los sistemas.

El término gestión de eventos se refiere al ciclo que se sigue al trabajar con esta información, desde la detección, pasando por su análisis, seguimiento, solución y cierre.

2.1 Términos clave

Antes de seguir profundizando en los sistemas SIEM, hay algunos términos que tienen que ser definidos, ya que es difícil encontrar definiciones precisas para muchos de estos términos. Las fuentes[ii] que se han utilizado son UIS DoD y CISSP.

Asset – Activo

Un activo es cualquier cosa dentro de la organización que tiene valor. En un nivel fácilmente cuantificable, esto incluye ordenadores, servidores, y equipamiento de red. Más allá, también se incluyen como activos las personas, los datos, los procesos, la propiedad intelectual, y la reputación.

Threat – Amenaza

Una amenaza es cualquier cosa con las capacidades y las intenciones de explotar una vulnerabilidad de un activo. El término amenaza es relativo, no es lo mismo una amenaza a un civil que a una organización. Las amenazas pueden clasificarse en dos categorías: estructuradas y desestructuradas.
Una amenaza estructurada utiliza tácticas formales y procedimientos, y tiene claramente definidos los objetivos. Normalmente se incluyen organizaciones criminales, grupos de hackers, agencias de inteligencia gubernamentales, y militares. Estos son típicamente grupos de individuos; aunque no es descartable que un solo individuo represente una amenaza estructurada.
Una amenaza desestructurada carece de motivación, habilidad, estrategia o la experiencia de una amenaza estructurada. Individuos aislados, o pequeños grupos dispersos frecuentemente representan una amenaza desestructurada. A menudo persiguen objetivos oportunistas, los cuales son elegidos porque aparentemente son vulnerables con facilidad.

Independientemente del alcance o la naturaleza de la amenaza, todas tienen algo en común: quieren robar algo de las organizaciones. Esto puede ser dinero, propiedad intelectual, reputación, o simplemente tiempo.

Vulnerability – Vulnerabilidad

Una vulnerabilidad es una debilidad del software, hardware o procedimental que puede permitir a un atacante la habilidad de manipular, ganar acceso o dañar cualquier tipo de activo. Esto puede ocurrir por código que no esté bien escrito y permita ataques de tipo buffer overflow, un puerto abierto en una red, o unas credenciales de acceso mal configuradas, hay que tener en cuenta que una persona también puede considerarse una vulnerabilidad.

Exploit

Un exploit es el método por el cual una vulnerabilidad es atacada. En el caso del software, esto puede ocurrir a través de un código que contiene un payload que permite al atacante realizar cualquier acción remota sobre un sistema, como abrir una ventana de comandos.

Risk – riesgo

Más adelante en este documento se hablará más sobre el riesgo, pero se puede definir como la medida de la posibilidad que una amenaza que puede explotar una vulnerabilidad. Sería ideal poder tener una métrica cuantificable del riesgo, pero esta tarea muchas veces es difícil de estimar dado las dificultades de valoración que tienen los activos en los sistemas de TI.

Anomaly – Anomalía

Es una ocurrencia observable en un sistema que está considerada fuera de lo normal. Las anomalías generan alertas utilizando herramientas de detección. Las anomalías pueden ser de algunos tipos como paquetes mal formados en la red, un contacto unusual con un host, una cantidad de datos elevada transmitidos en un corto periodo de tiempo.

Incident – Incidente

Un incidente es una violación o una amenaza inminente de violación de políticas de seguridad, de políticas de uso, o prácticas de seguridad. Simplificando podemos decir que un incidente es cuando algo malo ha ocurrido, está ocurriendo actualmente. Los incidentes pueden ir desde una instalación de malware, phising en el correo, un ataque de DoS. Un incidente está formado por eventos, pero la mayoría de los eventos no representan ningún incidente.

2.2 Función primaria de un SIEM

La función primaria de un SIEM es actual como un “recolector” central que pueda sintetizar los datos de seguridad sin procesar de las fuentes de información.

Importancia de la automatización:

- Reducir errores
- Menor tiempo de proceso
- Buscar menos complejidad

La clave de cada control que se implementa es la automatización[iii]. Cuando son los individuos los que tienen que supervisar a menudo se cometen fallos, haciendo vulnerables los sistemas que están encargados de proteger.
Uno de los principios básicos de seguridad es que, si hay una vulnerabilidad en el sistema, entonces esa vulnerabilidad tiene que repararse, entonces deja de representar un riesgo. Con los controles implementados de forma manual, las debilidades son las tareas humanas que interpretan y responden a las amenazas. A medida que se vaya mejorando el sistema, se pueden ir quitando las tareas que realizan los individuos, que son las debilidades del sistema. Esto no significa que las máquinas sean infalibles, sino que las tareas que realizarán los individuos serán mejorar los controles para aumentar su eficacia.

2.3 ¿Cuándo se necesita un SIEM?

Preguntas que se tienen que plantear antes de tomar la decisión de implementar un SIEM:

- ¿Es posible extraer información del 100% de las fuentes existentes en la organización?
- ¿Es segura la comunicación de la información de los log?
- ¿Los informes por defecto son suficientes? ¿es posible crear informes personalizados de manera sencilla?
- ¿Es posible establecer alarmas sobre cualquier cosa en los logs?
- ¿Qué herramientas de automatización existen?
- ¿Es posible auditar quiénes y qué logs se han revisado?
- ¿Las búsquedas de información específica son sencillas?
- ¿Se puede contextualizar y relacionar la información de los log para realizar tareas forenses?
- ¿Se puede probar que las políticas de seguridad, el control de cambios, el control de accesos están en uso y se están cumpliendo?
- ¿Se puede compartir de forma segura la información de los log con otras aplicaciones y usuarios?
- ¿Por qué controles empezar?
- ¿Cómo conseguir los mejores resultados en la implementación?

Priorizar los controles de sistemas en uso, implementar un hub central para procesar y correlacionar la información de las herramientas de seguridad.

2.4 Consejos para elegir controles

Los principales consejos para elegir los controles nos ayudarán a fijar las necesidades y los requisitos del sistema que necesita una organización:

- Las defensas deben centrarse en los ataques más comunes y más dañinos que ocurren actualmente, así como anticipar los que puedan ocurrir en un futuro próximo.
- Los entornos empresariales tienen que tener controles consistentes a través de todos los sistemas, para ser eficaces en la detección de los ataques.
- La automatización de las defensas tiene que ser probada y medida, en la medida de lo posible utilizando sistemas de medida automáticos.
- Para abordar los ataques actuales es necesario utilizar distintas técnicas para generar una defensa más consistente.

Cuando estos sistemas están bien configurados, tiene la capacidad de llegar a ser la “columna vertebral” de una red, recopilando y procesando fuentes de información, además de cumplir los requisitos de tratamiento y automatización. Lo que permite a las organizaciones securizar sus redes desde la capa física a la capa de aplicación permanentemente (24x7) .

2.5 Ventajas

Las ventajas de un SIEM son muchas, se han sintetizado las más importantes, que son las que están presentes en la mayoría de implementaciones disponibles de estos sistemas.

- Centralización de la información de seguridad; estos sistemas recopilan información del resto de aplicaciones y dispositivos. Lo que proporciona un punto de referencia a la hora de tratar esta información.
- Automatización de tareas, especialmente la recopilación de información puede consumir muchos recursos, la realización de estas de forma automática supone un ahorro importante de tiempos y costes.
- Seguimiento de los eventos, estos sistemas sirven para detectar anomalías de seguridad, y a su vez ver a medida que se efectúen actuaciones, puede verse su resultado.
- Evolución de la seguridad, al guardarse datos históricos de los sistemas que se monitorizan puede verse como varía la información y los controles de seguridad a lo largo del tiempo.
- Mejor manejo del riesgo, Independientemente del modelo de gestión de riesgos que se utilicen en las organizaciones, siempre contarán con dos elementos comunes: vulnerabilidades y amenazas. Los sistemas SIEM pueden mostrar si existen vulnerabilidades y si están siendo utilizadas por amenazas, tanto externas como internas.
- Métricas de seguridad, Cada control es capaz de convertirse en una métrica, de tal modo que se puede medir su funcionamiento de forma efectiva, esto nos indica si se está cumpliendo el objetivo de dicho control. De esta forma se facilita un proceso de mejora contínua y ayuda

a los propietarios del negocio establecer umbrales para los niveles de riesgos que son aceptables dentro de la organización.

3. LOGS

Un mensaje de log o registro, es lo que genera un sistema de computadora, dispositivo, software, etc. en respuesta a algún tipo de evento. El término logs o registros realmente se usa para indicar una colección de mensajes que se usarán colectivamente para “pintar una imagen” de alguna ocurrencia.
Esta es la base de información[iv] de todos los sistemas SIEM, estos logs van a ser utilizados para extraer información útil, analizando y categorizando. Cabe señalar que a veces el tipo de información puede clasificarse en más de una categoría. Los mensajes de inicio de sesión y de salida de usuario son relevantes, tanto para la gestión del inventario de usuarios, como para las políticas de seguridad de las cuentas.

Pero obtener la información de los log cuesta tiempo y trabajo, y a primera vista puede parecer una tarea abrumadora: el volumen total de datos puede ser desalentador. Hay que establecer una estrategia general para manejar los registros, en función de los tipos y formatos de registro. El objetivo de utilizar diferentes tipos y formatos de registro es doble. En primer lugar, familiarizarse con los mensajes de registro y los datos. En segundo lugar, establecer una visión general de los registros básicos para poder identificar y tratar con datos de registro nuevos o no vistos anteriormente, y abstraer relaciones entre ellos. Es una realidad que diferentes proveedores implementarán mensajes de registro en diferentes formatos, pero se utilizarán herramientas para manejar y administrar los datos de diversas fuentes. Cuanto más rápido se pueda comprender e integrar nuevos datos de log en el SIEM, más rápido se comenzará a obtener valor de él.

Los eventos dependen en gran medida del origen del mensaje de registro. Por ejemplo, los sistemas Unix tendrán mensajes de inicio de sesión y de cierre de sesión, los firewalls tendrán ACL aceptar y denegar mensajes, los sistemas de almacenamiento en disco generarán mensajes de registro cuando ocurran fallos o, en algunos casos, cuando el sistema perciba un fallo inminente.
Los datos de registro son el significado intrínseco que tiene un mensaje. O dicho de otra manera, los datos de registro son la información extraída de un mensaje para indicar la causa que generó el evento. Por ejemplo, un servidor web a menudo registra cada vez que alguien accede a un recurso (imagen, archivo, etc.) en una página web. Si el usuario que accede a la página debe autenticarse, el mensaje de registro contendrá el nombre del usuario. Este es un ejemplo de datos de registro: puede usar el nombre de usuario para determinar quién accedió a un recurso.

Los mensajes de registro se pueden clasificar en las siguientes categorías generales[v]: informativos, depuración, aviso, error y alerta.

Informativos: Los mensajes de tipo informativo están diseñados para que los usuarios y administradores sepan que ha ocurrido algo benigno. Por ejemplo,

Cisco IOS generará mensajes cuando se reinicie el sistema. Se debe tener cuidado, sin embargo. Si un reinicio, por ejemplo, ocurre fuera del mantenimiento normal o del horario comercial, es posible haya motivos para alarmarse.

Depuración: Los mensajes de depuración, generalmente se producen a partir de sistemas de software para ayudar a los desarrolladores a solucionar problemas, e identificar problemas con la ejecución del código de la aplicación. Conviene proteger estos mensajes, ya que los atacantes pueden utilizarlos para buscar vulnerabilidades en los programas.

Advertencia: Los mensajes de advertencia se refieren a situaciones en las que pueden faltar cosas, o ser necesarias para un sistema, pero la ausencia de las mismas no afectará el funcionamiento del mismo. Por ejemplo, si a un programa no se le proporciona la cantidad adecuada de argumentos de línea de comando, pero aún puede ejecutarse sin ellos, es algo que el programa podría registrar solo como una advertencia para el usuario o el operador. Esta información también puede servir para detectar comportamientos sospechosos, ya que puede relacionarse con otros eventos.

Error: Estos mensajes se utilizan para transmitir errores que ocurren en varios niveles en un sistema informático. Por ejemplo, un sistema operativo puede generar un registro de errores cuando no puede sincronizar los búferes en el disco. Desafortunadamente, muchos mensajes de error solo te dan un punto de partida sobre por qué ocurrieron. A menudo se requiere investigación adicional para llegar a la causa raíz del error.

Alerta: Una alerta está destinada a indicar que ha sucedido algo que necesita atención. Las alertas, en general, son el dominio de los dispositivos de seguridad y los sistemas relacionados con la seguridad, pero esta no es una regla rígida. Un Sistema de Prevención de Intrusiones (IPS) puede examinar todo el tráfico entrante, y aplicará una regla sobre si se permite o no una conexión de red determinada, en función del contenido de los datos del paquete. Si el IPS encuentra una conexión que podría ser maliciosa, puede realizar acciones preconfiguradas, y esto se registrará en un log como alerta.

3.1 Cómo se transmiten los mensajes de log

La transmisión y recolección de datos de log[vi], desde un ordenador o dispositivo conectado a la red, puede implementar un subsistema de log, por el cual puede generar un mensaje cada vez que determina que necesita hacerlo. Cada dispositivo tiene sus reglas para generar los mensajes y algunos son parametrizables. Estos pueden almacenarse localmente, pero su tratamiento sería manual y más costoso, en los sistemas de recolección de logs, se definirá un lugar donde se envíe y recopile el mensaje. Este lugar generalmente se conoce como loghost. Un loghost es un sistema informático, generalmente un sistema Unix o servidor de Windows, donde los mensajes de registro se recopilan en una base de datos de registros.

La forma más común de transmisión es a través del protocolo Syslog. El protocolo Syslog es un estándar para el intercambio de mensajes de log. Se encuentra comúnmente en sistemas Unix, pero existe para Windows y otras plataformas que no están basadas en Unix. Pero básicamente hay un componente de cliente y otro servidor, implementado sobre el Protocolo de Datagramas de Usuario (UDP), o sobre el Protocolo de Control de Transmisión (TCP). La parte del cliente es el dispositivo o sistema informático real que genera y envía mensajes de log. El lado del servidor normalmente se encontraría en un servidor de recopilación de log. Su trabajo principal es recibir los mensajes de registro basados en Syslog y almacenarlos, donde posteriormente se pueden analizar, realizar copias de seguridad y mantener un histórico para un uso a largo plazo.

Syslog no es el único mecanismo para la transmisión y recopilación de datos de log. Por ejemplo, Microsoft implementa su propio sistema de registro para Windows. Se llama registro de eventos de Windows. Cosas como el inicio y cierre de sesión de usuarios, mensajes de aplicaciones, etc. se almacenan en un formato de almacenamiento propietario. Hay aplicaciones de código abierto y comerciales que se ejecutan en la parte superior del Registro de Eventos de Windows que convertirá las entradas Syslog, desde donde se pueden reenviar a un servidor Syslog.

El protocolo simple de administración de red (SNMP) es un protocolo basado en estándares para administrar dispositivos en red. El protocolo se basa en dos conceptos: traps y polls (“avisos” y “sondeos”. Un trap es simplemente una forma de mensaje de registro que un dispositivo emite cada vez que algo ha sucedido. Se envía un trap a un recolector, que es análogo a un loghost. La figura del recolector se usa para administrar sistemas basados en SNMP. El sondeo es donde el rol de recolector puede usar el protocolo SNMP, para consultar a un dispositivo algunas variables predefinidas tales como estadísticas de interfaz, bytes transferidos dentro y fuera de una interfaz, etc. La diferencia clave entre SNMP y Syslog, es que se supone que SNMP debe estar estructurado con respecto al formato de datos. Pero esto no siempre se es así.
Las bases de datos también son el almacenen de los mensajes de log. En lugar de generar un mensaje Syslog, una aplicación puede escribir sus mensajes de registro en un esquema de base de datos. O en algunos casos, el servidor Syslog mismo puede escribir directamente en una base de datos relacional. Esto tiene grandes ventajas, especialmente en torno a proporcionar una forma estructurada de almacenar, analizar e informar sobre los mensajes de registro.

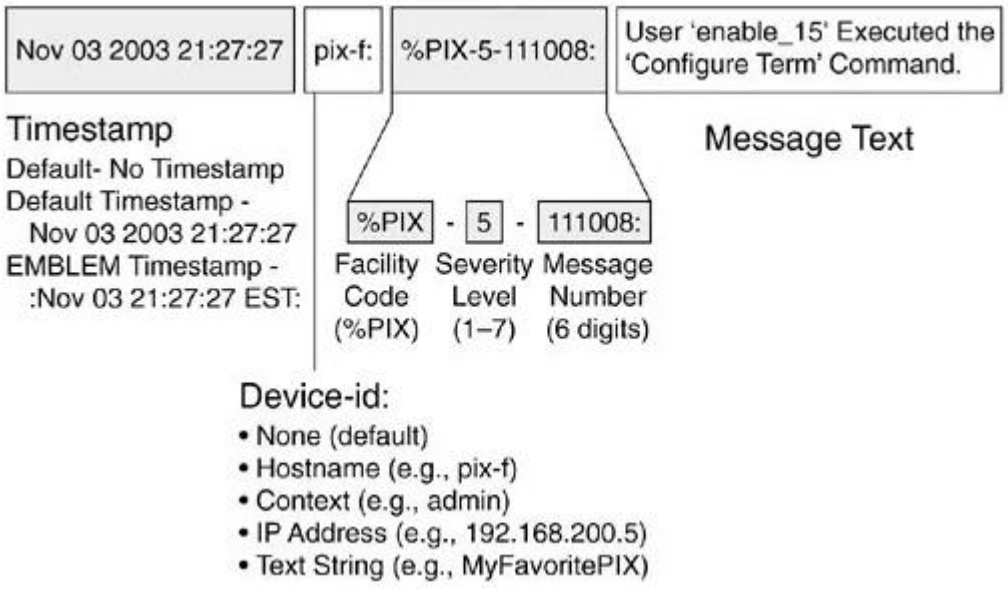
También hay formatos de log propietarios. Diversos fabricantes implementan sus propios mecanismos para generar y recuperar mensajes de log. En este ámbito, el proveedor suele proporcionar una Interfaz de programación de aplicaciones (API) en forma de bibliotecas C o Java, para que se puedan implementar interfaces de comunicación con otras herramientas. Algunos de ellos son:

- Log Extraction API (LEA): es la API de Checkpoint
- Security Device Event Exchange (SDEE) es el protocolo basado en el Lenguaje de marcado extensible (XML) de Cisco

- E-Streamer es el protocolo propietario de Sourcefire.

3.2 ¿Qué es un mensaje de Log?

Un mensaje de log está formado por tres partes básicas: una marca de tiempo (timestamp), la fuente o información de origen (source), y la información que se transmite (data). La marca de tiempo indica la hora a la que se generó el mensaje. La fuente es el sistema que generó el mensaje. Esto se representa típicamente en la dirección IP o el nombre del dispositivo. Finalmente, la información es el contenido semántico del mensaje. Lamentablemente, no hay un formato estándar de cómo se representan los datos en un mensaje. Algunos de los elementos de datos más comunes que encontrará en un mensaje de registro incluyen direcciones IP de origen y destino, puertos de origen y destino, nombres de usuario, nombres de programas, objetos de recursos (como archivos, directorios, etc.), bytes transferidos, etc.
La forma exacta en que se representa un mensaje de registro depende del sistema en origen. Como mencionamos anteriormente, Syslog es el formato más común utilizado por dispositivos y sistemas informáticos, este es un ejemplo de mensaje:



Il·lustració 1 - Mensaje de Log

Imagen obtenida de: https://stackify.com/syslog-101/

En la parte inicial del mensaje se ve la marca de tiempo, en la que se indica mes, día y año. En la parte central, los datos que identifican el origen del mensaje, puede ser el hostname, la dirección IP, un texto, etc. En la parte final aparece la información propiamente dicha del mensaje.

3.3 Infravaloración de los Log

Los logs son poco apreciados en muchos entornos empresariales. A menudo, los registros se ignoran por completo, y solo se miran cuando el espacio en disco es bajo. En ese momento, generalmente se eliminan sin revisión. Y en algunos casos, algunos de los mensajes en los registros podrían haber indicado por qué el disco estaba lleno.
Muchos fabricantes no quieren se usen los logs. Los vendedores de herramientas de administración de TI le dirán que necesita sus productos, con pequeños agentes en cada host que informan sobre la misma información se encuentra en los registros de log. Pero, por supuesto, si puede obtener la información de sus registros, no se necesitan sus productos.
El análisis de logs no es fácil y requiere cierto trabajo. Los registros vienen en una variedad de formas y tamaños, y en ocasiones puede ser difícil extraer información de ellos. Además, puede haber una gran cantidad de datos para trabajar. Por ejemplo, algunos sitios recopilan varios gigabytes por semana de datos de logs, y tal volumen puede parecer abrumador para empezar.

El análisis de logs efectivo también requiere conocimiento amplio de su entorno y sus sistemas; hay que saber qué es bueno y qué es malo, qué es sospechoso y qué es normal. Lo que es malo o extraño para un sistema puede ser perfectamente normal para otro. Por ejemplo, un usuario que inicia sesión en un país extranjero puede ser una amenaza, especialmente si todos sus usuarios son locales. Sin embargo, otro sitio puede tener usuarios repartidos en todas partes, por lo que es difícil para ellos determinar qué se consideraría inusual.

En parte esta es la razón por la que no hay herramientas de análisis de logs que no necesiten configuración, porque es cada entorno y sus políticas lo que determina qué desea obtener de sus logs.

A pesar de esta infravaloración, los logs pueden indicar muchas cosas sobre lo que está sucediendo en un sistema, desde la información de rendimiento hasta la detección de fallos y la detección de intrusos. Los logs pueden ser una buena fuente de información "forense" para determinar "qué sucedió" después de un incidente. Y los logs pueden hacer un seguimiento de auditoría. A lo largo de este TFM desarrollaremos estas y otras funcionalidades de los logs dentro de las herramientas SIEM.

3.4 Herramientas de análisis y recopilación de Logs

Siguiendo una evolución tecnológica, a medida que fué aumentando el tamaño de los sistemas, empezó a desarrollarse un mercado de herramientas gestión y análisis de logs, que ha ido creciendo significativamente a lo largo de los años. Muchas herramientas, como grep y awk, están integradas en multitud de los servidores y dispositivos de red. Existen soluciones tanto de código abierto como propietario, y una serie de herramientas para la centralización, para administrar e informar sobre los distintos logs dentro de su entorno. Estas herramientas van desde la búsqueda básica básica, la recopilación centralizada de logs, y las más complejas que incorporan informes específicos de cumplimiento y alertas en tiempo real.

Herramientas básicas de análisis

Existen varias herramientas simples pero potentes para realizar análisis de Logs, algunas integradas en los sistemas operativos Linux / Unix, otras disponibles para su descarga de forma gratuita a través de Internet, y algunas otras forman parte de paquetes de software. Revisaremos algunos de ellas[vii].

Grep es una utilidad de línea de comandos que se encuentra hoy en la mayoría de los sistemas operativos Linux y Unix, y que busca en archivos de entrada utilizando un patrón o expresión regular.
Dentro de estos archivos se pueden buscar fácilmente con grep para encontrar información útil para revisiones diarias de logs y análisis forense. Grep es muy dependiente del usuario que tiene que conocer el término de búsqueda o información de lo que está buscando

Referencias: también se ha portado a muchos otros sistemas operativos, incluido Windows con licencia GNU
(http://gnuwin32.sourceforge.net/packages/grep.htm) y como parte de Cygwin (http://www.cygwin.com/).

AWK es otra herramienta disponible inicialmente en plataformas Linux / Unix, posteriormente se extendió a otros sistemas operativos; Awk se puede utilizar como una herramienta de análisis de logs, sirve para extraer y aislar información de los archivos que reciba como entrada.

Referencias: se puede encontrar más información sobre awk en: http://www.gnu.org/software/gawk/manual/gawk.html

Microsoft Log Parser. Windows tiene sus propios registros de log y formatos específicos. Una utilidad gratuita disponible de Microsoft se llama Log Parser. La actualización más reciente de esta utilidad se puede descargar desde el Centro de descarga de Microsoft en http://www.microsoft.com/en-us/download/details.aspx?id=24659 .
Log Parser es una herramienta que proporciona acceso de consulta a datos basados en texto, como archivos de registro, XML y CSV, así como fuentes de datos del sistema operativo Windows, como el Visor de Eventos, el Registro, el sistema de archivos y el Directorio Activo. Una de las funciones más útiles de Log Parser es la capacidad de consultar el registro de eventos de Windows mediante sentencias similares a SQL.

Otras herramientas básicas a considerar

Las herramientas anteriores permiten una gran parametrización, pero también hay que considerar otras más sencillas que son utilizadas muy frecuentemente:

- Tail : sirve para revisar el final del archivo de logs o el final de cualquier fichero.
- Head : lo contrario que tail, sirve para recuperar la parte superior de un archivo de logs grande, y así hacerlo más manejable.

- Sed : es una utilidad de análisis sintáctico como awk, es útil para hacer búsquedas y reemplazar textos, para que los logs sean más claro de leer, o para formatear la salida de un fichero y hacerlo legible para otras utilidades.
- Logwatch : es una utilidad para analizar y revisar tus logs sin conexión, tiene una interfaz conectable que permite alto grado de personalización.
- Lire : es un conjunto de utilidades que permiten generar informes personalizados a partir de archivos de log. Es similar a Microsoft Log Analyzer, se pueden filtrar y generar informes de un archivo de logs. Es compatible con una variedad de archivos de log de aplicaciones y software fuera de la plataforma Windows.

Hay muchas maneras de utilizar estas herramientas básicas, además están disponibles libremente, o muy posiblemente ya se estén utilizando dentro de muchas de organizaciones. Se pueden usar para buscar y refinar datos de logs, y de esta forma aislar un ataque. Sin embargo, estas herramientas están mejor adaptadas para organizaciones pequeñas, y para el análisis y revisión de logs de forma manual. Por otra parte estas herramientas tienen algunas limitaciones, y hay que considerar los requisitos de la organización, así como los tiempos y recursos necesarios. Lo que nos lleva a buscar formas de agrupar la información.

3.5 Utilidades para centralizar la información de logs

Cuando se ha visto la cantidad de información disponible en los ficheros de logs, ha surgido la necesidad de centralizar la información de distintos ficheros, para permitir relacionar los registros entre ellos, y así facilitar el análisis y su recopilación. Hay una serie de herramientas disponibles gratuitamente para ayudar a centralizar la información de logs. Algunas son: syslog, syslog-ng, rsyslog y Snare.

Syslog trata de ser una solución de registro ubicua en todas las plataformas. Existe soporte integrado para el registros de syslog en dispositivos de red, sistemas Unix / Linux, así como también muchas extensiones de bibliotecas de lenguajes de programación, para facilitar a los desarrolladores de aplicaciones utilizar los registros de Syslog. Hay dos componentes principales, el cliente Syslog que existe en el sistema que genera los logs y se puede configurar para enviar sus registros localmente o a un servidor Syslog centralizado. Y el servidor, en el que se encuentra el daemon syslog, que recibe logs de otros clientes syslog configurados para enviar datos. A parte de la centralización de la información de logs, existen otras cualidades no disponibles en Syslog y que pueden ser necesarias en las organizaciones:

- Entrega garantizada de los mensajes: en muchas de las implementaciones de syslog, se comunican y transmiten los mensajes a través de UDP, por el puerto 514. En redes ocupadas, el tráfico del protocolo UDP puede perder mensajes para permitir que se entreguen otros mensajes de alta prioridad, especialmente del protocolo TCP. Por lo tanto, no hay garantía de que cada mensaje de log de las fuentes se registre en el servidor Syslog centralizado.

- Transporte seguro de mensajes : Aunque se propuso en RFC-5425 (http://tools.ietf.org/html/rfc5425), pocas implementaciones de Syslog cifran los mensajes de log para que no se puedan leer o se alteren en una red. Esto se está volviendo más importante para las organizaciones, ya que aumentan el número de sitios remotos y utilizan hosts virtualizados en la nube, y la información que se transmite es confidencial.
- Mantenimiento de los datos de origen : En algunos entornos se puede utilizar un servidor de retransmisión, para recopilar registros de un sitio remoto, o segmento de red, y posteriormente reenviar los registros al servidor Syslog central. En muchas implementaciones de Syslog, la información de origen sobre dónde se generó originalmente el registro se pierde como parte del proceso de retransmisión. Esto puede hacer que sea casi imposible determinar de dónde vino un mensaje.

Rsyslog. Es otra opción de código abierto y también está disponible de forma gratuita en muchas distribuciones de Linux. Existen algunas diferencias de características entre syslog-ng y rsyslog, pero en general, rsyslog y syslog-ng son bastante compatibles en el conjunto de funcionalidades. Algunos beneficios al elegir rsyslog vs. syslog-ng son los siguientes:

- Rsyslog es el registrador predeterminado para los sistemas Redhat y Fedora Linux. Rsyslog será una opción más frecuente para organizaciones con una gran base de equipos Redhat.
- Rsyslog recientemente agregó soporte para Hadoop (HDFS). Las organizaciones que utilizan Hadoop, para mejorar su inteligencia de seguridad y la extracción de logs, podrán integrar rsyslog con esta función, con una necesidad limitada de escribir scripts personalizados para cargar datos en Hadoop.

Snare. Windows y su entorno registran una cantidad de información valiosa en el registro de eventos de Windows. Desafortunadamente, este registro de eventos es una tecnología de propietaria de Microsoft sin soporte nativo para la mensajería de estilo Syslog. Hay opciones disponibles en el mercado para recuperar la información del registro de eventos de los sistemas Windows, y agregar estos datos con Syslog-ng y Rsyslog. Sin embargo, hay otra alternativa gratuita, Snare (http://www.intersectalliance.com/projects/index.html) permite que los eventos de Windows se envíen a través de syslog. Snare tiene sus propias opciones de recopilación de agentes y servidores.

3.6 Herramientas de análisis de Logs

Vistas algunas de las herramientas básicas para la recopilación de información. Nos encontramos con la necesidad de un análisis en profundidad de los datos de los logs, o la necesidad de generar alertas sobre eventos en tiempo real, así como generar informes, o la incorporación de automatización, que reducen la cantidad de tiempo necesaria para realizar las revisiones de registro diarias, y mejorar el análisis forense cuando ocurren eventos. Para estas necesidades hay otro tipo de herramientas como las que se comentarán a continuación.

La herramienta OSSEC[viii] es una herramienta de código abierto para la retención y el análisis de logs. Se incluyen agentes para muchas plataformas y sistemas, además de soporte para recibir logs de fuentes de Syslog existentes. Incluye soporte para una opción sin agente, que puede verificar la integridad de archivos en muchas plataformas donde no se puede instalar el agente OSSEC. Tiene soporte para sistemas host VMWare para monitorear infraestructura de virtualización. Otra de las principales ventajas es que reduce el proceso manual de análisis de registros, tiene reglas preinstaladas para ayudar a usar alertas en tiempo real, y una interfaz de usuario basada en la Web. OSSEC es un proceso ligero, requiere pocos recursos. El el servidor de administración centralizado solo está disponible para sistemas Unix / Linux. En 2009, Trend Micro adquirió OSSEC y ha seguido manteniendo la herramienta abierta y gratuita; también ofrece soporte comercial. Entre las principales características podemos encontrar:

- Flexibilidad en la parametrización para la retención de logs.
- Posibilidad de crear reglas y alertas en función de las necesidades.

OSSIM[ix]. Un ejemplo de agrupar las funcionalidades de varias, es más que administración y retención de registros, esta herramienta pertenece a un grupo de productos que se considera un sistema de Gestión de Eventos e Información Segura (SIEM). Ya que proporciona lo que se denominan las "5 capacidades de seguridad esenciales": descubrimiento avanzado, evaluación de vulnerabilidades, detección de amenazas, monitorización del comportamiento e inteligencia de seguridad. Las siguientes herramientas de código abierto se utilizan internamente para proporcionar estas capacidades:

- Arpwatch se utiliza para la detección de anomalías de direcciones MAC.
- P0f se utiliza para la detección pasiva y el análisis de cambios en sistemas operativos.
- Pads se utiliza para la detección de anomalías de servicio
- Nessus se utiliza para la evaluación de vulnerabilidad y para la correlación cruzada (sistema de detección de intrusos (IDS) frente a Vulnerability Scanner)
- Snort se usa como un sistema de detección de intrusos (IDS), y también se usa para la correlación cruzada con Nessus
- Tcptrack se utiliza para obtener información de datos de sesión que pueden recopilarse para la identificación de ataques.
- Ntop crea una impresionante base de datos de información de red para la detección de anomalías de comportamiento.
- Nagios para supervisar la información de disponibilidad del servicio y del host en función de una base de datos de activos.
- Osiris como un sistema de detección de intrusiones basado en host (HIDS)
- Snare un recolector de logs para sistemas de Windows.
- OSSEC como IDS basado en host

OSSIM también incluye sus propias herramientas, incluido un motor de correlación genérico con compatibilidad con directivas lógicas e integración de registros con complementos.

OSSIM incluye un conjunto bastante grande de tablas y vistas de datos, formado un sistema completo de administración de información de seguridad, con un solo sistema de administración y retención de Logs.

Otras herramientas

Además de las vistas anteriormente, se expondrán a continuación alguna herramienta más:

- Logsurfer : es una herramienta útil para revisar registros en tiempo real. Otro beneficio es que intenta agrupar mensajes que relacionados en un único evento lógico.
- LogHound : es una herramienta de investigación, intenta encontrar patrones de línea frecuentes en archivos de logs utilizando un algoritmo breadth-first (búsqueda en anchura).
- Log2Timeline : es útil para realizar análisis forenses. La herramienta extrae información marca de tiempo, de algunas fuentes no

tradicionales, como la papelera de reciclaje, el historial de Firefox, los datos del servidor proxy, etc. Esto puede ser útil cuando existe la preocupación de que los datos o los registros se hayan modificado o alterado a lo largo del tiempo.

Splunk [x]: es una oferta comercial que ofrece características similares a OSSEC. Splunk le permitirá centralizar logs para análisis forense y correlación, así como generar alertas en tiempo real, estas pueden ser basadas en los tipos de eventos, que requieren mayor investigación o acción. Algunas diferencias clave con otras herramientas es la amplia variedad de fuentes de logs que admite, los paneles de control en tiempo real para revisar la actividad, informes personalizables y cuadros de mandos, APIs compatibles con proveedores para ayudar a su integración en la infraestructura de seguridad. Existen versión versiones comerciales y gratuitas con algunas limitaciones.

NetIQ Sentinel[xi], al igual que OSSIM, es otro producto que entra en la categoría de ser un SIEM. Sentinel incluye la detección de anomalías y la información de gestión de identidades, utilizando fuentes adicionales de datos, que pueden ser útiles en el manejo de la respuesta a incidentes, así como la realización de análisis forenses sobre los eventos. También existen distintas versiones comerciales y gratuitas.

IBM q1Labs[xii] QRadar : es la solución de gestión de logs de IBM. Tiene un amplio soporte de orígenes de logs, y una amplia variedad de funciones de búsqueda e informes. QRadar se puede adquirir con algunas de las líneas de productos adicionales, incluida una oferta SIEM completa para organizaciones. Uno de los principales diferenciadores para QRadar son sus informes adaptado para muchos de los marcos de cumplimiento normativo actuales, dando un respaldo importante a los procesos de auditoría de:

- Payment Card Industry Data Security Standard (PCI DSS)
- North American Electric Reliability Corporation (NERC)
- Health Insurance Portability and Accountability Act (HIPAA)
- Sarbanes-Oxley (SOX), and Federal Information Security Management Act (FISMA)

Loggly[xiii] es un proveedor recopilación de logs en la nube. El producto admite muchas de las fuentes de datos, así como la publicación de datos a través de una API REST programable, y la publicación de datos a través de HTTP y HTTPS. Una diferencia clave respecto al resto de sistemas mencionados, será que no es necesario instalar hardware adicional en su entorno para retener sus registros, así como que se externaliza el mantenimiento y actualización del sistema de logs.

La evolución de este tipo de herramientas ha llevado a generar un nuevo concepto llamado SIEM, en el que se agrupan las características más básicas de recolección de información y automatización, junto con las más avanzadas de análisis y gestión de la información.

4. Controles críticos y SIEM

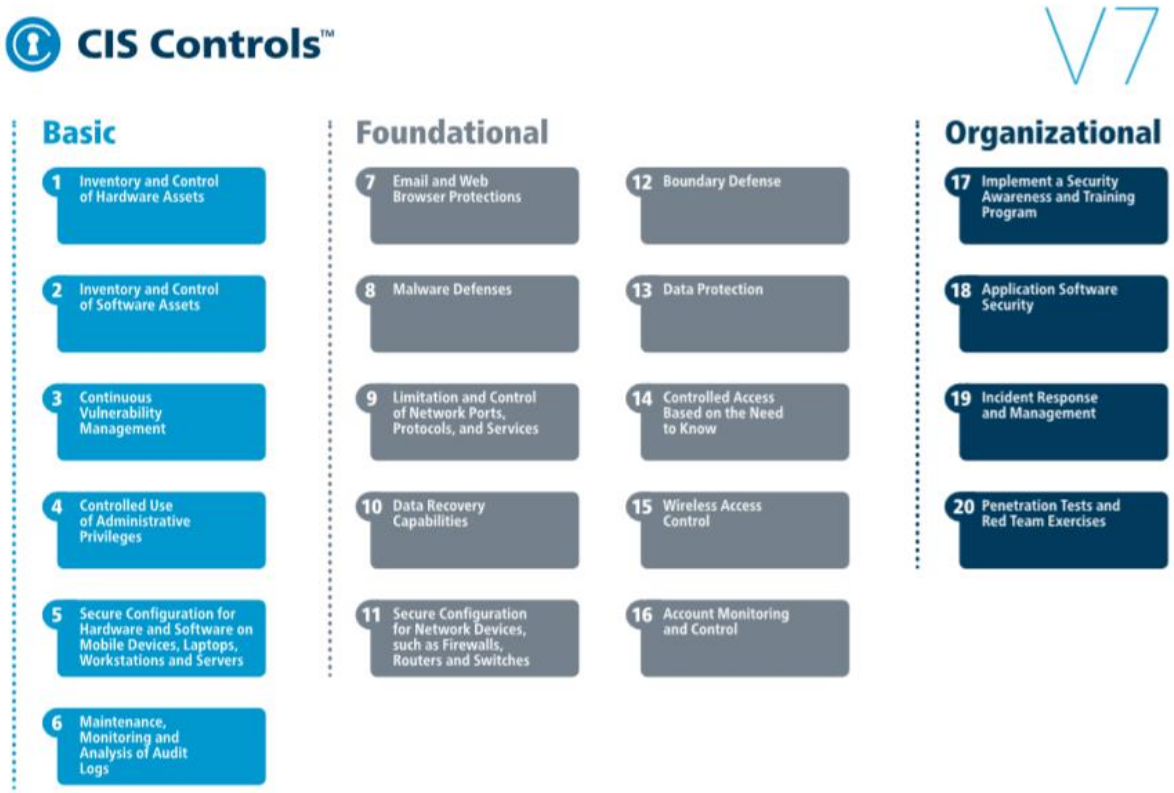
Los controles[xiv] van a ser el objeto de automatización sobre la información recogida en el SIEM. Los siguientes controles críticos, sirven para tener una visión de conjunto de las funcionalidades del sistema.

Taula 1 - Controles Críticos

Control Crítico	Relación con herramientas SIEM
1: Inventario de dispositivos autorizados y bloqueados	El sistema SIEM debería utilizarse como la base de datos de inventario de los activos autorizados. El conocimiento de esta información puede permitir la detección de amenazas a través de la localización, regulaciones vigentes, criticidad de datos.
2: Inventario de software autorizado y bloqueado	Igual que en el control 1, el sistema SIEM debería utilizarse como la base de datos de inventario de software para la correlación con la actividad de red de las distintas aplicaciones.
3: Evaluación de vulnerabilidades de forma continua y parcheado.	Los sistemas SIEM pueden relacionar las vulnerabilidades en el contexto de actividad actual del sistema, para determinar si las vulnerabilidades están siendo explotadas.
4: Uso controlado de de permisos administrativos.	Cuando los principios de este control no se cumplen, por ejemplo un navegador web ejecutado por un administrador, los sistemas SIEM pueden relacionar estos eventos para detectar violaciones y generar alertas.
5: Configuraciones seguras para hardware y software en portátiles, sobremesa y servidores.	Las vulnerabilidades conocidas siguen siendo una vía principal para los ataques exitosos. Si un dispositivo tiene un parámetro de red mal configurado se detecta durante un escaneo, esa mala configuración se informa al SIEM. Esto ayuda a resolver incidentes
6: Mantenimiento, monitorización y análisis de logs de auditoría.	Los sistemas SIEM son el núcleo del motor de análisis, que tratan los eventos en tiempo real, recogiendo y centralizando logs críticos.
7: Protección del correo electrónico y los navegadores	Las vulnerabilidades de los navegadores y los clientes son las más frecuentes, es necesario el registro de sus eventos en el SIEM para su análisis.
8: Defensas de malware	Los objetivos del malware son muy amplios y su forma de actuación muy dinámica, estas defensas tiene que estar integradas con el SIEM.
9: Limitación y control de puertos, protocolos y servicios de red.	Como resultado de los escaneos (CCE), si un sistema tiene algún puerto abierto, protocolo o servicio se tiene que reportar al SIEM, ya que se puede monitorizar el tráfico sobre esos puertos prohibidos, protocolos o servicios. Esta regla se puede utilizar para conocer qué

Control Crítico	Relación con herramientas SIEM
	puertos, protocolos o servicios son útiles y cuáles no.
10: Capacidad de recuperación de datos.	Los procesos y herramientas utilizados para respaldar adecuadamente la información crítica, utilizando una metodología comprobada para la recuperación óptima de la misma.
11: Configuraciones seguras para dispositivos de red como firewalls, routers y switches.	Cualquier error en la configuración de los dispositivos de red se reportará al SIEM para un análisis consolidado.
12: Defensa perimetral.	Las violaciones de red se reputarán al SIEM, para su correlación con el inventario de elementos autorizados en el repositorio del sistema SIEM.
13: Prevención de pérdida de datos	las violaciones de las reglas de pérdida de datos, como los eventos detectados en la monitorización (CCE), deben reportarse al SIEM, para relacionar estos eventos con el inventario y otras informaciones sobre actividad para detectar brechas de seguridad.
14: Acceso controlados basados en la necesidad de saber	Los sistemas SIEM pueden relacionar la actividad de los usuarios, con sus permisos y roles para detectar violaciones, comprobando los privilegios.
15: Control de dispositivos wireless	Las configuraciones incorrectas de dispositivos y las intrusiones inalámbricas se deben reportar al SIEM, para la relación de incidencias.
16: Monitorización y control de cuentas.	La actividad anormal sólo puede ser detectada cuando se compara a una línea base de actividad buena conocida. Esta línea base para cumplir este control es elaborada por el sistema SIEM, y a medida que se van generando nuevas líneas base, pueden irse comparando con la línea base aprobada por el sistema.
17: Programa de formación y concienciación de la seguridad	Para todos los roles de la organización, enfocándose en su conocimiento y habilidades, para adaptarse a las necesidades cambiantes de la seguridad.
18: Aplicación de seguridad en software	Cuando se realiza un análisis de configuraciones (CCE), las vulnerabilidades que son descubiertas en aplicaciones también deben reportarse a un punto central. Estas pueden relacionarse con otros eventos.
19: Gestión y respuesta de incidentes	Proteger la información de la organización desarrollando e implementando una infraestructura de respuesta a incidentes, para descubrir rápidamente un ataque y contener el daño de forma efectiva, erradicando la presencia del atacante.

Control Crítico	Relación con herramientas SIEM
20: Test de penetración y Red Team Exercises	Pruebe la fortaleza general de la defensa de una organización (la tecnología, los procesos y las personas) simulando los objetivos y las acciones de un atacante.



Il·lustració 2 - Controles Críticos y Clasificación

Los controles se pueden organizar en tres categorias[xv] Basic, Foundationaly Organizational.

5. Productos comerciales

Según Gartner en el año 2017, las tres principales herramientas SIEM son Splunk, LogRhythm y QRadar de IBM

.

Il·lustració 3 - Gartner 2017 Magic Quadrant for SIEM

Gartner 2017 Magic Quadrant for SIEM
https://www.splunk.com/en_us/form/gartner-siem-magic-quadrant.html

6. Desarrollar, comprar o externalizar

La herramienta adecuada para cada organización depende los requisitos a cumplir, sobre todo en aquellas empresas en las que la seguridad no forma parte de su portfolio comercial, y por lo tanto no se puede extraer un beneficio directamente de su implantación. El aumento de los incidentes de seguridad, ha hecho que muchas organizaciones se den cuenta del coste potencial, y de la posible pérdida de negocio, al no elegir el sistema correcto o al no tener suficientes recursos para monitorizar los sistemas. Aunque una organización no sepa qué papel desempeña en la seguridad, la gestión efectiva de logs, podemos suponer que poder revisar, correlacionar y analizar logs, son tareas fundamentales para identificar los sistemas y clientes afectados, por lo que las organizaciones podrán cerrar brechas en su seguridad y continuar con su actividad. Este enfoque nos permite introducir los sistemas SIEM en las organizaciones, para que la dirección comprenda que son una pieza más de los sistemas IT críticos, dentro de la arquitectura. Con esta visión se podrá evaluar las decisión de comprar, construir o externalizar la solución[xvi]. En muchos casos, pueden darse combinaciones de comprar y construir, comprar y subcontratar, etc., puede proporcionar mejores beneficios generales que un enfoque único para satisfacer las necesidades de su organización. A medida que analicemos las diversas herramientas a través de este capítulo, debe considerar una serie de ventajas y desventajas en la decisión de comprar, construir o subcontratar.

6.1 Desarrollar una solución

Algunas organizaciones deciden crear su propia solución para la administración de logs, debido a la disponibilidad de soluciones de código abierto, así como el deseo de personalizar y ajustar soluciones específicas para su entorno y necesidades. En muchos casos, este enfoque permite que una organización comenzar un desarrollo, con recursos existentes y costes reducidos. Una serie de ventajas y riesgos están asociados con este enfoque:

Ventajas:

- Es probable que obtenga un sistema y una solución que sean exactamente lo que desea para su entorno
- Puede hacer cosas que no se pueden encontrar en soluciones comerciales o de código abierto, ya que en muchos casos puede modificar y actualizar el código en el sistema.
- Puede elegir y diseñar la plataforma, las herramientas y los procesos para su sistema.
- No se requieren costos iniciales para adquirir el sistema.
- Este tipo de proyectos motivan al equipo de trabajo. ¡Es divertido!

Riesgos:

- Al ser el propietario del sistema, deberá asignar recursos y tiempo para mantenimiento del sistema, así como para cualquier actualización necesaria para continuar cumpliendo los requisitos.

- Sin soporte de terceros, las personas del departamento de IT ¡Son el personal de apoyo!
- Si alguno de los miembros clave del personal que construyó el sistema abandona la empresa¿Se podrá contratar, retener y capacitar al personal para continuar manteniendo el sistema?
- A medida que la organización crezca ¿El sistema podrá manejar el volumen de logs de todos los sistemas y continuar escalando?

6.2 Comprar una solución

Muchas organizaciones pueden considerar que construir un sistema de administración de logs consume demasiado tiempo, y pueden no tener los recursos para dedicar a la construcción, incluyendo el mantenimiento de estos sistemas, sobre todo si el desarrollo no es una competencia central de la organización. Las organizaciones más grandes también necesitan acuerdos de soporte con los proveedores, para garantizar el tiempo de actividad y los requisitos legales. Estos acuerdos de soporte y los requisitos legales a menudo no se pueden cumplir con soluciones propias basadas en productos de código abierto, sobre todo si tienen que ser certificados por terceras empresas. Debido a estos y muchos otros factores, muchas organizaciones buscan comprar directamente estos sistemas. Las siguiente ventajas y riesgos deben tenerse en cuenta al comprar un sistema:

Ventajas

- "Cash and carry": pague y obtenga una solución para la administración y análisis de logs.
- Las soluciones comerciales admiten una amplia variedad de fuentes y formatos de registro.
- Los acuerdos de soporte generalmente incluyen acuerdos de servicio, para acordar el tiempo de actividad, y el tiempo de respuesta ante problemas.
- Actualizaciones y mejoras del producto, para adaptarse a los requisitos y/o la normativa.
- Posibilidad de contratar servicios de instalación, así como soporte in-situ para ayudar al despliegue del sistema, y capacitar al personal interno.

Riesgos

- A parte del costo inicial del sistema, se obtendrá un sistema en poco tiempo, pero se necesitará contratar o capacitar al personal para configurarlo y usarlo. Las organizaciones deben considerar el impacto económico fuera del coste inicial del sistema
- Si se adquiere el sistema ¿Cuenta la organización con personal que tenga las habilidades para aprender, usar y obtener valor del sistema?
- Puede haber lagunas en el sistema, que no tienen soporte para aplicaciones instaladas en su entorno, o procesos específicos para sus requisitos.
- Madurez del producto, así como su longevidad, son indicadores para que la organización pueda cambiar de proveedor en el futuro si es necesario.

6.3 Externalizar

Otras organizaciones consideran que la externalización es la mejor alternativa, especialmente en áreas donde no tienen recursos para construir u operar y mantener las soluciones que adquieren. Al igual que la compra, la externalización permite a las organizaciones cumplir con los requisitos, tener soporte, y alinearse con las necesidades legales del entorno. Las siguiente ventajas y riesgos deben tenerse en cuenta a la hora de externalizar un sistema:

Ventajas:

- Alguien externo se preocupará por las tareas diarias y los requisitos de su organización. Esto libera recursos internos para enfocarse en otros elementos básicos para el negocio.
- La externalización nos permite reducir la infraestructura, ya que el proveedor de outsourcing albergará la infraestructura de la organización.
- Menos personal necesario para dedicarlo a la administración de logs, y las actividades diarias de revisión y análisis.

Riesgos:

- Alguien externo se está preocupando de los problemas y puede no tener el mismo conocimiento y las necesidades que el personal interno.
- Puede haber lagunas en el sistema y no tener soporte para aplicaciones instaladas en su entorno.
- La organización pierde el control de sus datos. Esto supone un riesgo adicional de pérdida si se alojan fuera de la organización, también puede dificultar el cambio de proveedores en el futuro.
- El volumen de datos de su organización puede ser enorme, por lo que puede afectar los SLA de su sistema impactando en el rendimiento.
- El acceso a los datos de registro puede estar limitado en función de las API proporcionadas por el proveedor, del mismo modo que puede haber limitaciones en los tiempos de retención de datos.

7. Arquitectura ELK

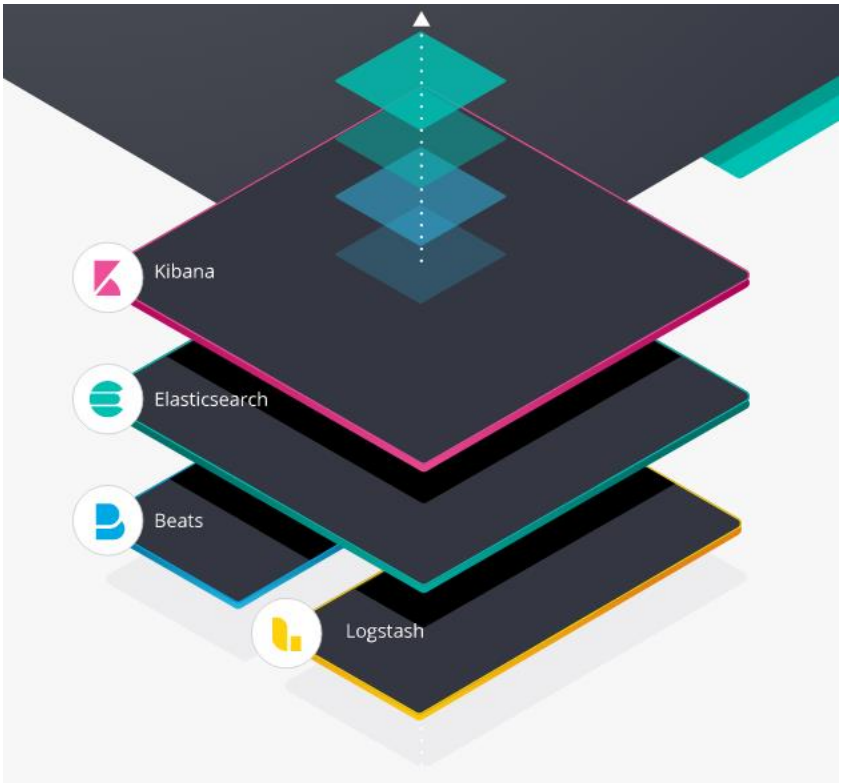
La arquitectura està basada en los siguientes Componentes:

Elasticsearch[xvii]: Es un motor de búsqueda basado en Lucene. Proporciona la búsqueda de texto completo, es distribuido, y gracias a su interfaz web junto a la capacidad de almacenamiento de documentos JSON, le proporciona gran versatilidad. Está desarrollado en Java, y su utilización es bajo licencia Apache. Hay múltiples clientes desarrollados en distintos lenguajes: Java, .NET (C #), PHP, Python, Apache Groovy, Ruby, etc.

Logstash[xviii]: Es una herramienta ETL, permite la extracción, transformación y carga de datos desde y hacia múltiples fuentes. Permite gestionar el flujo de la información, aplicando distintos tipos de filtros, así como construir estructuras

de datos a partir de las entradas. Siendo muy util la función de enriquecimiento de la información entre distintas fuentes.

Kibana[xix]: Es una utilidad de visualización de datos open source, proporciona capacidades de extracción de información sobre contenido indexado de Elasticsearch. Permite la utilización de múltiples tipos de gráficos: barras, líneas, dispersión, circulares, información geográfica, etc. Permitiendo de forma sencilla el manejo de grandes volúmenes de información.



Il·lustració 4 - Arquitectura ELK

El sistema propuesto con la arquitectura de ELK tiene múltiples aplicaciones comerciales, y al ser muy flexible permite su configuración para utilizarse como SIEM.
Los componentes que lo forman se comunican vía REST o HTTP, lo que permite su interconexión con otras muchas aplicaciones y desarrollos a medida.
Para el TFM se ha implementado dentro de una máquina virtual con Linux Ubuntu 18.04 LTS, con 4GB de RAM, 20 GB de HD, 4 CPUs y 128 MB de vídeo.

Il·lustració 5 - Flujo de información ELK

Esquema de la arquitectura: https://logz.io/learn/complete-guide-elk-stack/

Como todos los elementos que forman parte del sistema se van a instalar sobre el mismo HOST, en los parámetros de configuración habrá que prestar especial atención a la configuración de los puertos de comunicación entre las aplicaciones:

Aplicación	**Puerto TCP**
Elasticsearch	9200
Kibana	5601
Logstash	5000

Hay distintas formas de instalación de este producto, para la configuración de la arquitectura en un único HOST y para gestionar manualmente los inicios/paradas de los Servicios, para más detalles ver el Anexo I - Instalación.

7.1 Watcher

El plugin Watcher[xx] para elasticsearc proporciona alertas y notificaciones basadas en cambios en los datos, permite la automatización de un amplio rango de casos de uso como por ejemplo monitorización de infraestructuras, trazar la actividad de red, monitorización de aplicaciones, etc.

Todos estos casos tienen propiedades comunes entre ellos, por ejemplo cambios en los datos que pueden identificarse a través de una consulta sobre elasticsearch, así resultados de consultas que pueden contrastarse sobre una condición. Cuando esta condición es cierta se pueden realizar una o más acciones, como enviar un correo o una notificación a otro sistema.

Un Watcher está formado por cuatro bloques principales:

- Schedule - se define el intervalo de tiempo en el cual se va a lanzar la consulta y comprobar la condición.

- Query - se especifica la consulta que se va a utilizar como datos de entrada para comprobar la condición.
- Condition - la condición determinará si se ejecutan las acciones o no, en función de su evaluación respecto a los datos de entrada.
- Actions - una o más acciones, como la notificación de correos o envio de datos a sistemas externos, se realizarán cuando se cumplan las condiciones del Watcher.

Se añaden Watchers a los sistemas para automatizar acciones cuando se cumplen ciertas condiciones. Estas condiciones están basadas en los datos que se cargan en el Watch, estos datos se denominan Payload. Este payload puede ser obtenido desde diferentes recursos.

Se adjunta el código de un Watch que busca errores en un log de eventos:

```
PUT _xpack/watcher/watch/log_errors
{
  "metadata" : { ❶
    "color" : "red"
  },
  "trigger" : { ❷
    "schedule" : {
      "interval" : "5m"
    }
  },
  "input" : { ❸
    "search" : {
      "request" : {
        "indices" : "log-events",
        "body" : {
          "size" : 0,
          "query" : { "match" : { "status" : "error" } }
        }
      }
    }
  },
  "condition" : { ❹
    "compare" : { "ctx.payload.hits.total" : { "gt" : 5 }}
  },
  "transform" : { ❺
    "search" : {
       "request" : {
         "indices" : "log-events",
         "body" : {
           "query" : { "match" : { "status" : "error" } }
         }
       }
    }
  },
  "actions" : { ❻
    "my_webhook" : {
```

```
    "webhook" : {
      "method" : "POST",
      "host" : "mylisteninghost",
      "port" : 9200,
      "path" : "/{{watch_id}}",
      "body" : "Encountered {{ctx.payload.hits.total}} errors"
    }
  },
  "email_administrator" : {
    "email" : {
      "to" : "sys.admino@host.domain",
      "subject" : "Encountered {{ctx.payload.hits.total}} errors",
      "body" : "Too many error in the system, see attached data",
      "attachments" : {
        "attached_data" : {
          "data" : {
            "format" : "json"
          }
        }
      },
      "priority" : "high"
    }
  }
 }
}
```

❶ Metadata - Opcionalmente se pueden adjuntar datos estáticos para el watch.
❷ Trigger - La programación temporal del trigger hace que el watch se ejecute cada 5 minutos.
❸ Input - Esta entrada busca errores en el índice log-events, y carga la respuesta en el watch payload.
❹ Condition - Esta condición comprueba si hay más de 5 errores en los el índice log-event.Si los hay, la ejecución continua para todas las acciones.
❺ Transform - Si se cumple la condición, esta transformación carga todos los errores en el payload, buscando los errores con el tipo de búsqueda: query_then_fetch. Todas las acciones del watch tienen acceso a este payload.
❻ Actions - Este watch tiene dos aciones. La primera acción se llama my_webhook notifica a un tercer sistema sobre el problema. La segunda acción se llama email_administrator envia un correo con prioridad alta al administrador del sistema. El watch payload que contiene los errores es adjuntado al correo.

Cuando se añade un watch, se registra un trigger en el apropiado trigger engine, de tal forma que un sheduler va iniciando los distintos watch que tenga registrados en función de los intervalos de tiempo que tengan programados cada uno.

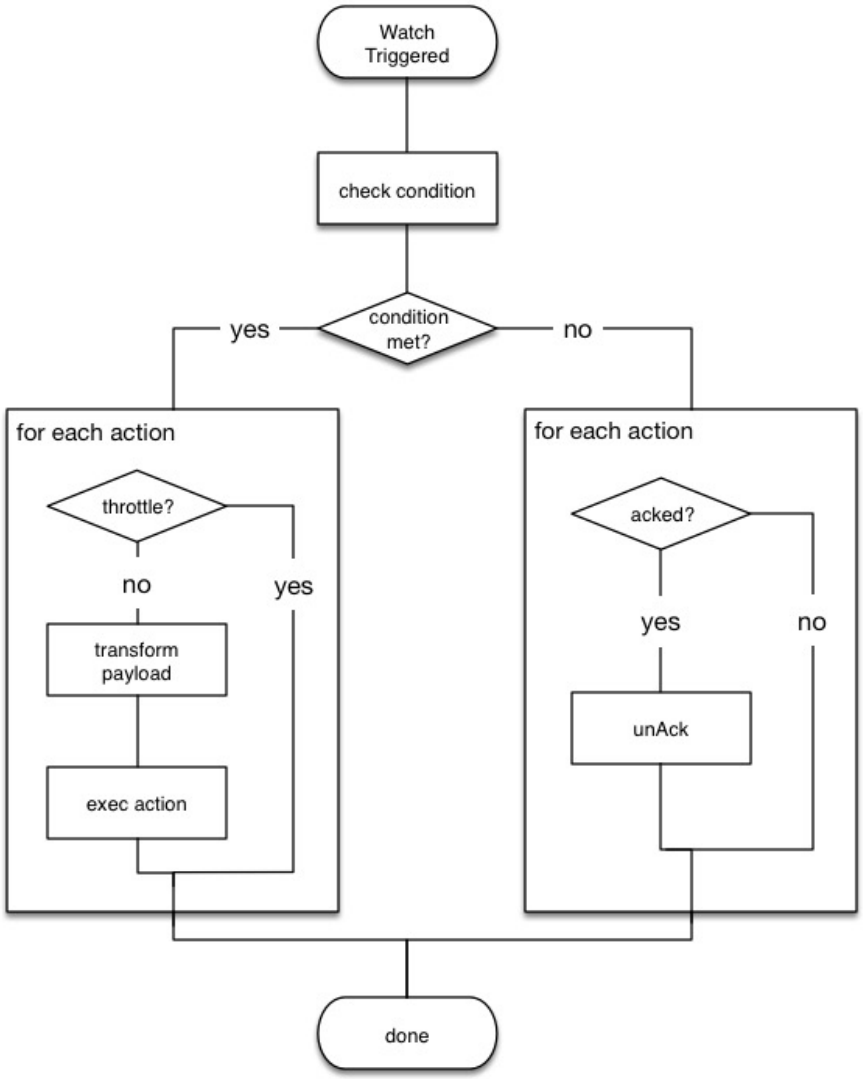
Si se detiene el servicio del Watcher, el sheduler se detiene también, y cuando se lanza la la ejecución este entra en una cola, se añade un registro en el historial de watch, y se le asigna el estado de awaits_execution.

Al comenzar la ejecución, se crea un contexto, en el cual se proporcionan las plantillas y los scripts, con acceso a los metadatos, al payload y a la información del trigger.

Durante la ejecución el Watch, con la información cargada en el payload ya es accesible para los pasos siguientes del proceso. La evaluación de la condición determina si se continúa o no con el proceso; si esta evaluación es verdadero (true), el proceso avanza a los siguientes pasos, si la evaluación es falso (false), se detiene la ejecución del Watch. Opcionalmente se aplican las transformaciones, si son necesarias, y se realizan las acciones.

Cuando termina la ejecución del Watch, el resultado es almacenado en el historial, la información que se guarda es el resultado de la condición, el tiempo de ejecución y el resultado de cada acción que ha sido ejecutada.

Diagrama de ejecución:



Il·lustració 6 - Diagrama de ejecución de un Watcher

Imagen URL: https://www.elastic.co/guide/en/x-pack/6.2/images/watch-execution.jpg

Referencia: https://www.elastic.co/guide/en/x-pack/6.2/how-watcher-works.html

8. Implementación de un SIEM

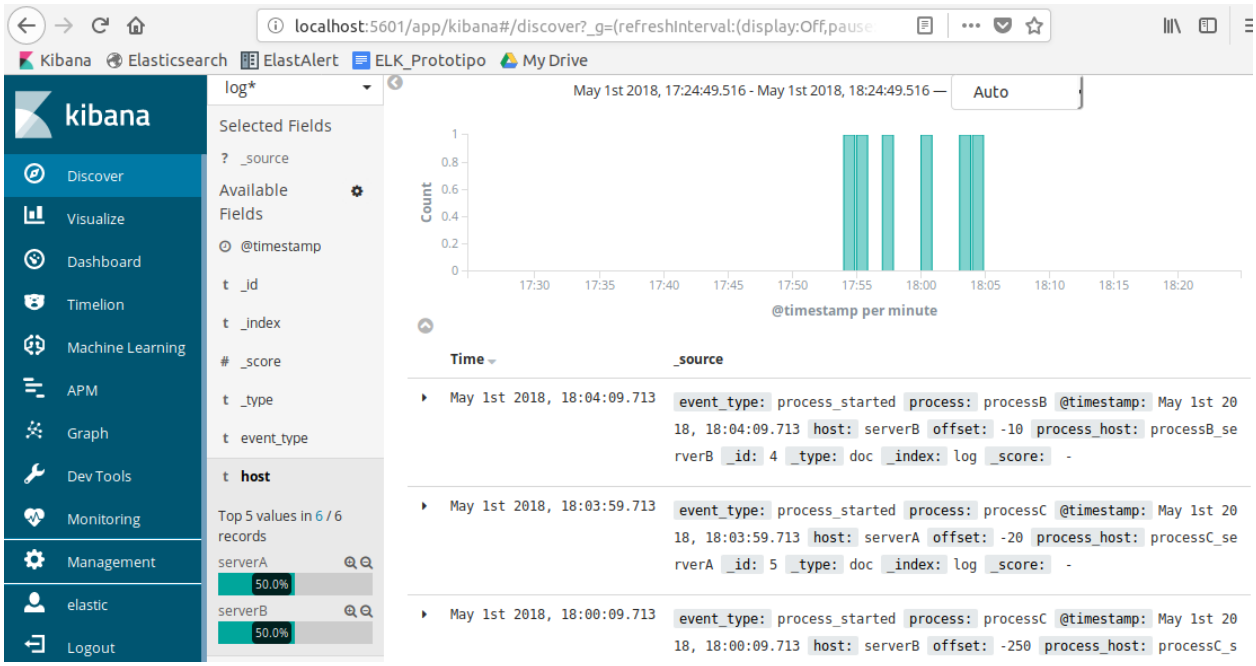
Se ha desarrollado un laboratorio sobre una màquina virtual, en la que se han simulado unas entradas de datos, se han puesto en marcha unas alertas (Watcher), y se han disparado acciones cuando se han cumplido ciertos criterios. Los controles implementados se describen en los siguientes apartados.

8.1 Alerta primera ejecución de un proceso

Este Watch tiene como objetivo alertar si un proceso se inicia en un servidor por primera vez.

El reloj examina los N minutos previos para los procesos iniciados. Esta lista se usa a su vez para buscar datos anteriores a N minutos, para ver si los procesos se han iniciado históricamente. Cualquier diferencia da como resultado una alerta.

En la siguiente imagen se puede ver un pantallazo de este ejercicio:



Il·lustració 7 - Alerta primera ejecución de un proceso

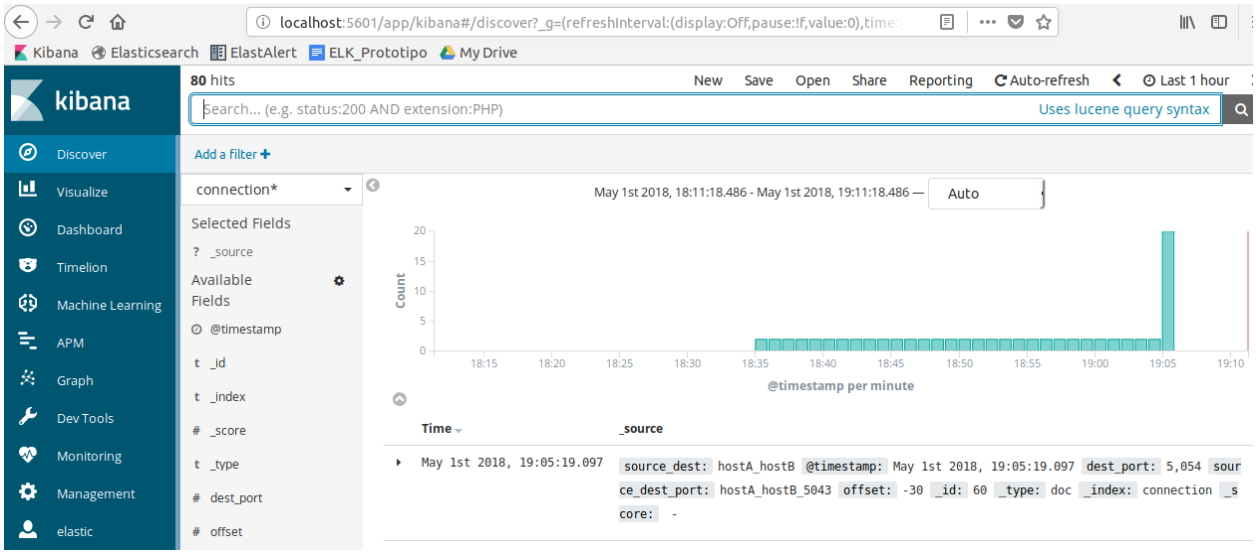
8.2 Alerta de escaneo de puertos

El objetivo de este Watcher es detectar y alertar si un servidor establece una gran cantidad de conexiones a un destino, a través de una gran cantidad de puertos.

Una escaneo de puertos se produce cuando se establecen una gran cantidad de conexiones entre dos servidores, en una gran cantidad de puertos distintos. Esto se puede detectar como una gran cantidad de documentos en elasticsearch, con valores de puerto únicos, para los mismos valores de origen-destino.

Esta alerta evita asociar un valor exacto a "alto". En su lugar, pretende basar la interpretación de alto en los datos disponibles y el comportamiento habitual. Además, esta alerta debería poder hacer frente a una gran cantidad de dispositivos> 100k.

En la siguiente imagen se puede ver un pantallazo de este ejercicio:



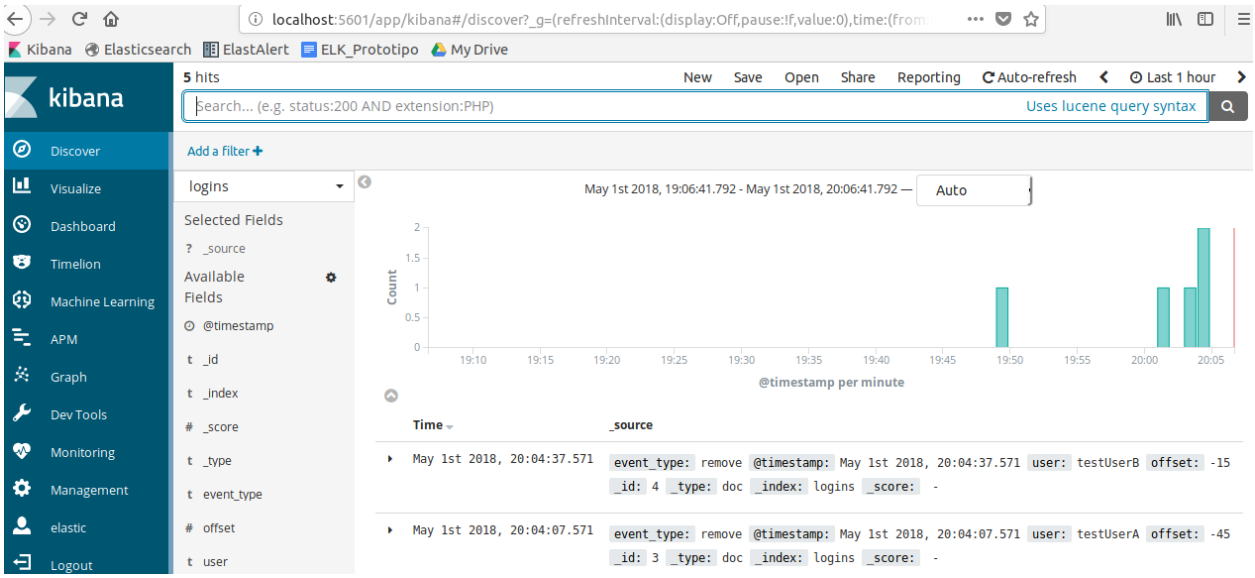
Il·lustració 8 – Alerta Escaneo de puertos

8.3 Alerta de actividad de cuenta sospechosa

Este Watcher trata de detectar y alertar si un usuario es creado en Active Directory / LDAP y posteriormente eliminado dentro de N minutos siguientes.

En un sistema comprometido un asaltante podria realitzar actividades de este tipo para manipular los registros que deja su actividad fraudulenta sobre el sistema.

En la siguiente imagen se puede ver un pantallazo de este ejercicio:

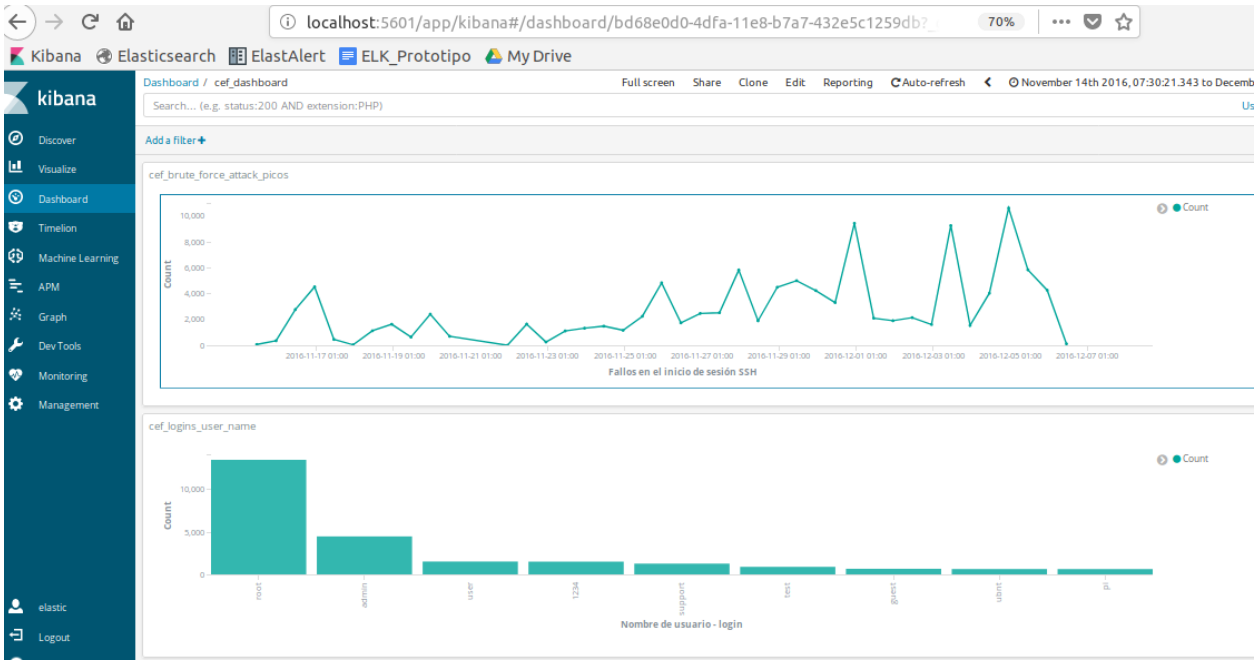


Il·lustració 9 - Alerta actividad de cuenta sospechosa

8.4 Alerta analizando SSH

En este ejemplo se van a utilizar datos en el formato CEF, para analizar los eventos de log en los inicios de sesión de un servicio de SSH.
La finalidad es detectar los logins correctos desde una dirección externa, y también detectar los ataques de fuerza bruta que puede recibir el sistema, definiendo esta situación como una secuencia de N intentos fallidos de login, seguidos de un login exitoso.

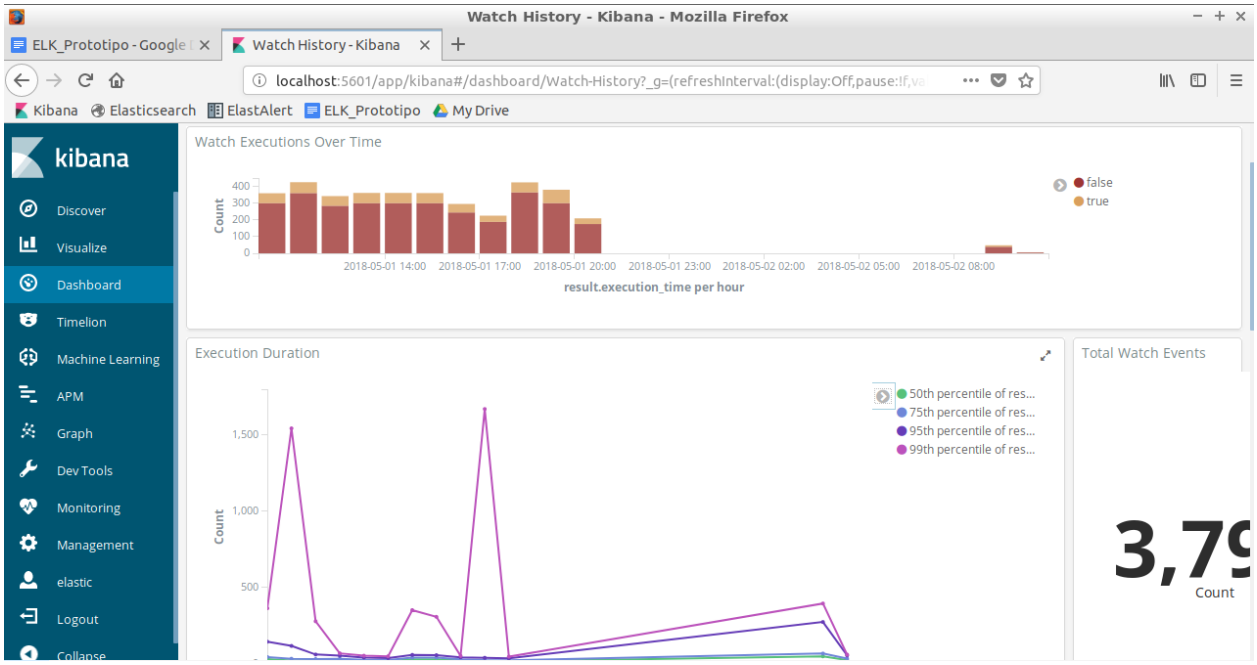
En la siguiente imagen se puede ver un pantallazo de este ejercicio:



Il·lustració 10 - Análisis de tráfico SSH

8.5 Cuadro de mando

Ver los resultados de forma gráfica ayuda mucho a comprender el estado del sistema que se está monitorizando, con esta finalidad se han agrupado los eventos que han lanzado los distintos Watcher.
El cuadro de mando se ha parametrizado utilizando uno de los ejemplos que se proporcionan con el plug-in X-pack, en el que se permite ver la historia de los Watcher que han estado en ejecución en el sistema.



Il·lustració 11 - Cuadro de mando SIEM

La primera gráfica representa los watcher que se han ejecutado en un periodo de tiempo, diferenciando en rojo aquellos de resultado negativo y en naranja aquellos de resultado positivo.

La segunda gráfica representa el tiempo de ejecución de los Watcher, nos sirve para monitorizar el rendimiento del sistema y su carga de trabajo.

Los resultados pueden ser llamativos, debido a que es un entorno de laboratorio en el que se han buscado ejemplos para que se lancen las alarmas de una forma rápida.

Referencia:
https://github.com/elastic/examples/tree/master/Alerting/watcher_dashboard

10. Conclusiones

Las principal lección que se estrae de la elaboración de este TFM es que desde se dispone abundante información en los sitemas y los aplicativos que están siendo utilizados, y es necesario dar valor a estos datos (logs, registros de eventos, trazas, ...) ampliando sus funcionalidades en el campo de la seguridad. Actualmente los log están infrautilizados.
Me gustaria destacar las posibilidades de automatización que dan este tipo de herramientas SIEM, ya que permiten agilizar los controles y adaptarlos al cicló de vida de los aplicativos y sistemes.

Los objetivos iniciales no se han conseguido al 100%, ya que desde el punto de vista teórico, me hubise gustado profundizar más en la evaluación del riesgo, ya que es la principal herramienta a la hora de decidir los controles a implementar. En la parte de desarrollo, inicalmente se habían elegido más controles, y las horas necesarias para la implementación de cada uno han sido mayores de las estimadas, por lo que para cumplir con la planificación se han reducido el número de controles implementados.

Se ha seguido la metodologia inicial con ciertos cambios, ya que los controles y los juegos de datos para las pruebas se han elegido para poder reproducirse con mayor facilidad, manteniendo el correcto funcionamiento del SIEM

Como lineas de Trabajo futuro se destacan dos aspectos:

- Cuadros de mando: este tipo de herramientas permiten tenir una vision en su conjunto del sistema, y se pueden enfocar a los requsitos que haya que alcanzar.
- Técnicas de Machine-learning, para explotar la información de los sistemas, ya que muchas veces se pueden aplicar estos modelos
- Integración con otras herramientas de Seguridad, como sistemas de dectección de intrusos.

11. Glosario

SIEM - Security Information and Event Management
APT - Advanced Persistent Threat
CCE - Common Configuration Enumeration
ERD - Entity Relationship Diagram
ROSI - Return on Security Investment
PITSR - Perceived IT Security Risk
SLA - System Level Agreement
DoS - Denial of Service
IPS - Intrusion detection system

12. Bibliografía

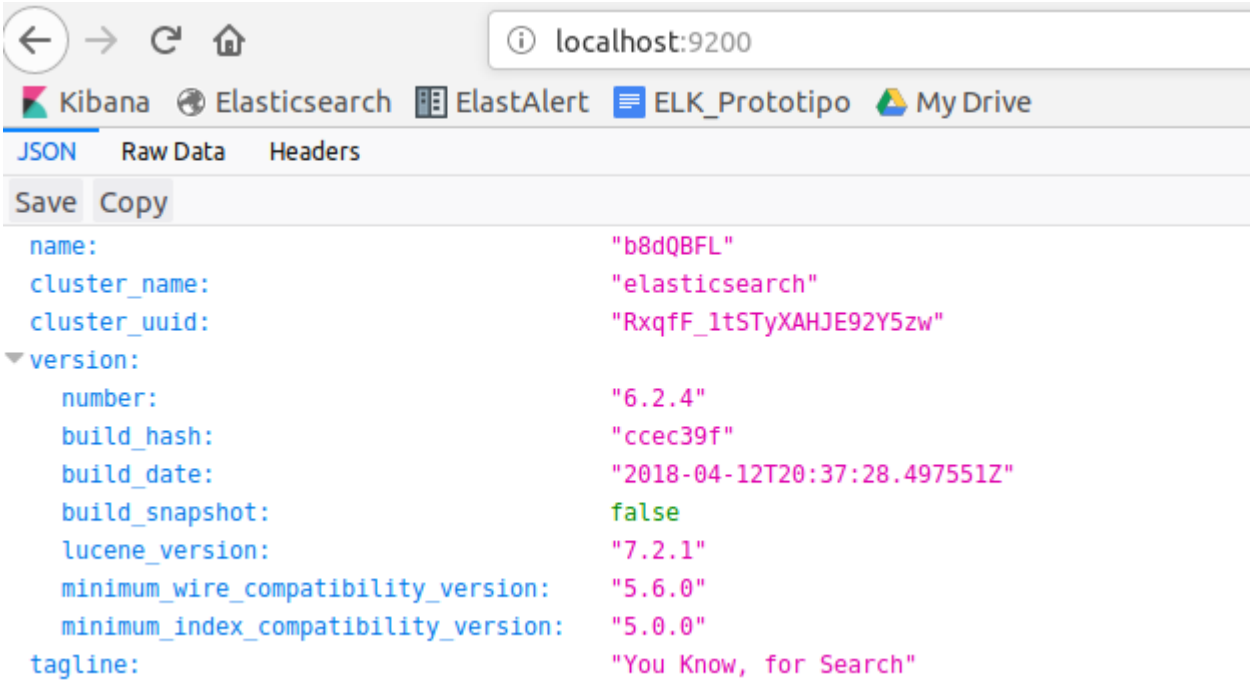
[i] Licencia Apache - https://en.wikipedia.org/wiki/Apache_License
[ii] Términos clave - Applied Network Security Monitoring: Collection, Detection & Analysis (ISBN - 978-0124172081)
[iii] SANS Institute InfoSec Reading Room: Implementing the 20 Critical Controls with Security Information and Event Management (SIEM) Systems https://www.sans.org/reading-room/whitepapers/analyst/implementing-20-critical-controls-security-information-event-management-siem-systems-34965
[iv] Logs, trees, forest - Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management ISBN 978-1597496353
[v] Categorías de los mensajes de Log - Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management ISBN 978-1597496353
[vi] How is Log Data Transmitted and Collected? - Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management - ISBN 978-1597496353
[vii] Basic tools for log analysis - Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management - ISBN 978-1597496353
[viii] Más información de la herramienta OSSEC en la web www.ossec.net
[ix] Más información de la herramienta OSSIM en la web http://www.alienvault.com/
[x] Web de Splunk www.splunk.com
[xi] NetIQ Sentinel https://www.netiq.com/products/sentinel
[xii] IBM q1Labs QRadar http://q1labs.com/products.aspx
[xiii] Loggly http://loggly.com/
[xiv] SANS Institute InfoSec Reading Room: Implementing the 20 Critical Controls with Security Information and Event Management (SIEM) Systems https://www.sans.org/reading-room/whitepapers/analyst/implementing-20-critical-controls-security-information-event-management-siem-systems-34965
[xv] Center for Internet Security: https://www.cisecurity.org/controls/
[xvi] Tools for Log Analysis - Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management ISBN 978-1597496353
[xvii] Elasticsearch - https://en.wikipedia.org/wiki/Elasticsearch
[xviii] Logstash - https://www.elastic.co/products/logstash
[xix] Kibana - https://en.wikipedia.org/wiki/Kibana
[xx] Plugin Watcher - https://www.elastic.co/guide/en/watcher/current/introduction.html

13. Anexo I – Instalación

13.1 Instalación de Elasticsearch

Elasticsearch va a ser el componente central de todo el sistema, el método de instalación elegido consta de tres pasos:

- Descargar el archivo .TAR, y descomprimirlo (Enlace de descarga: https://www.elastic.co/downloads/elasticsearch)
- Ir a la carpeta bin, y ejecutar el comando elasticsearch para iniciar el proceso
- Probar que ha ido bien el proceso y que es accesible elasticsearch, se puede utilizar el comando "curl http://localhost:9200/" o abrir una ventana de un navegador y probar la dirección http://localhost:9200/



Il·lustració 12 - Datos instalación ElasticSearch

Configuración de elasticsearch

El fichero de configuración se llama config/elasticsearch.yml y contiene parámetros importantes como los datos de identificación del nodo y del cluster, rutas de almacenamiento de datos y logs de actividad, configuración de la memoria para optimizar el rendimiento, datos de red y puerto de escucha.
La configuración por defecto permite que se inicien los servicios con un nodo funcional, por lo que no vamos a modificar ninguno de estos parámetros, solamente añadiremos los datos de la cuenta de correo electrónico de Gmail utilizada para enviar las alertas, al final del fichero las líneas añadidas son las siguientes:

```
# ========================== Elasticsearch Configuration =============
#
# ---------------------------------- Various -----------------------------------
#
# Configuración de correo para la notificación de alertas

xpack.notification.email.account:
  work:
        profile: gmail
        email_defaults:
        from: mi.elk.siem@gmail.com
        smtp:
        auth: true
        starttls.enable: true
        host: smtp.gmail.com
        port: 587
        user: mi.elk.siem@gmail.com
        password: <poner aquí la clave de la cuenta de correo>
```

X-pack plugin

Se ha decidido utilizar este plug-in ya proporciona una rápida configuración de los roles de usuario, y la funcionalidad de alertas Watcher va a ser de gran utilizadad para el SIEM implementado. Los pasos para la instalación son los siguientes:

- Ejecutar ./bin/elasticsearch-plugin install x-pack
- Arrancar elasticsearch
- Configurar passwords ./bin/x-pack/setup-passwords interactive

```
elk@elk-VirtualBox:~/Downloads/elasticsearch-6.2.3$ ./bin/x-pack/setup-passwords
interactive
Initiating the setup of passwords for reserved users elastic,kibana,logstash_system.
You will be prompted to enter passwords as the process progresses.
Please confirm that you would like to continue [y/N]y


Enter password for [elastic]:
Reenter password for [elastic]:
Enter password for [kibana]:
Reenter password for [kibana]:
Enter password for [logstash_system]:
Reenter password for [logstash_system]:
Changed password for user [kibana]
Changed password for user [logstash_system]
Changed password for user [elastic]
elk@elk-VirtualBox:~/Downloads/elasticsearch-6.2.3$
```

En este último paso se han configurado los usuarios del sistema: elastic , kibana, logstash

13.2 Instalación de Kibana

Kibana es la herramienta de visualización, entre los tipos de existentes, la instalación elegida lleva los siguientes pasos:

- Descargar y descomprimir Kibana (fichero TAR): https://www.elastic.co/downloads/kibana
- Abrir config/kibana.yml en un editor
- Asignar la variable elasticsearch.url para que apunte a la instacia de Elasticsearch que se ha instalado en el punto anterior

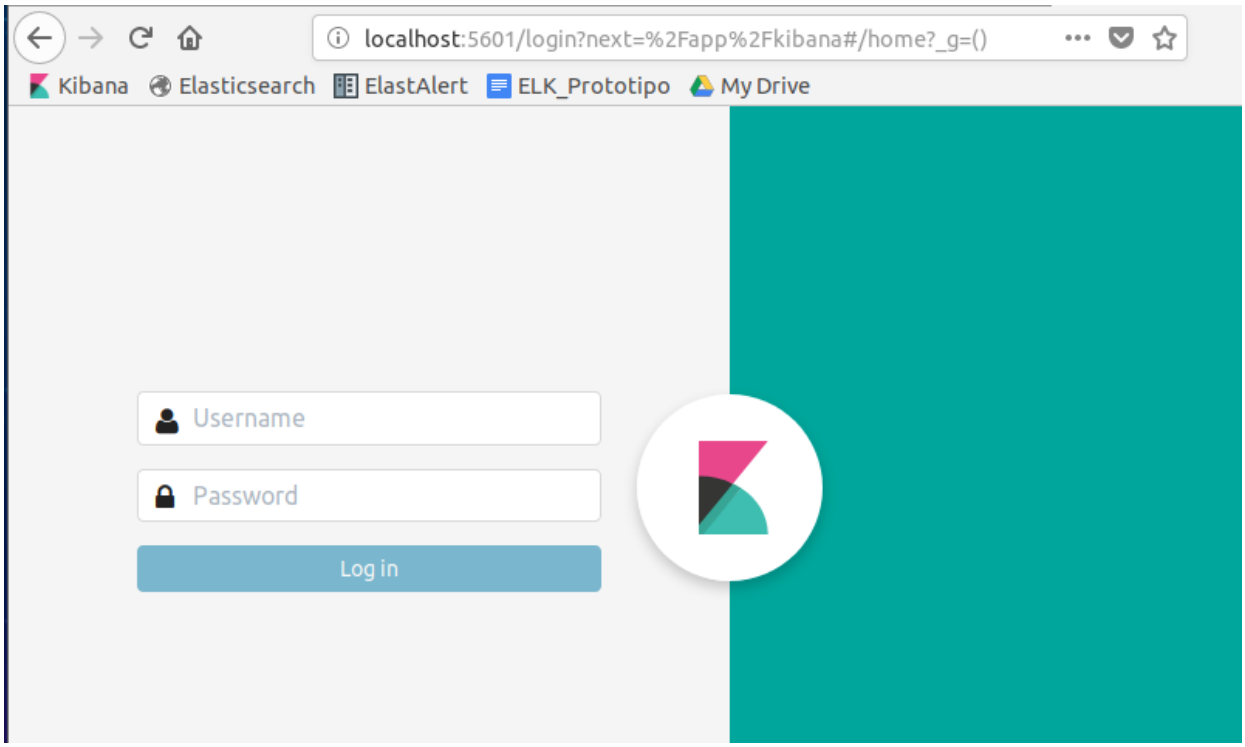
```
# Kibana is served by a back end server. This setting specifies the port to use.
#server.port: 5601

# The URL of the Elasticsearch instance to use for all your queries.
elasticsearch.url: "http://localhost:9200"

# If your Elasticsearch is protected with basic authentication, these settings provide
# the username and password that the Kibana server uses to perform maintenance on the Kibana
# index at startup. Your Kibana users still need to authenticate with Elasticsearch, which
# is proxied through the Kibana server.
#elasticsearch.username: "user"
#elasticsearch.password: "pass"
elasticsearch.username: "elastic"
elasticsearch.password: "poner aqui la clave del usuario"
```

En este fichero de configuración se pueden especificar los datos de red, datos de parametrización de logs, tiempos de respuesta, configuración de seguridad y certificados,

- Ejecutamos el siguiente comando bin/kibana
- Probamos el acceso http://localhost:5601

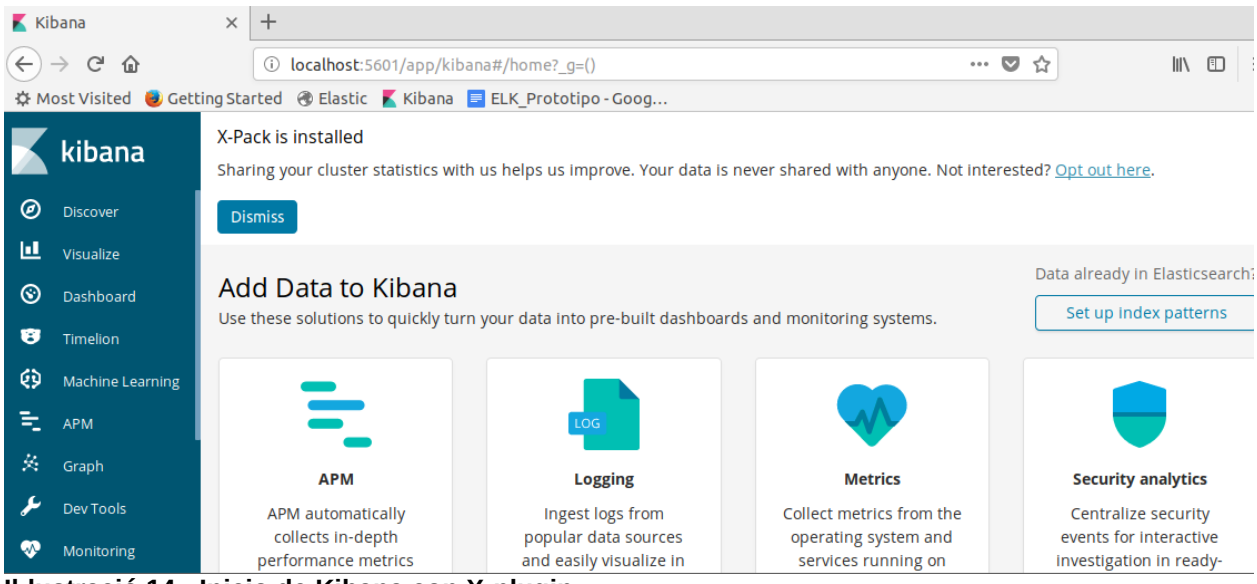


Il·lustració 13 - Inicio de Kibana

Configuramos el X-plugin

- Ejecutar ./bin/kibana-plugin install x-pack

La siguiente vez que entremos a Kibana veremos el siguiente panel:



Il·lustració 14 - Inicio de Kibana con X-plugin

Referencia: https://www.elastic.co/start

13.3 Instalar Logstash

Logshtash nos permite la extracción, transformación y la carga de datos en Elasticsearch, entre las opciones de instalación se ha elegido la siguiente:

- Descargar y descomprimir el fichero TAR, https://www.elastic.co/downloads/logstash
- Configurar el fichero logstash.conf, indicaremos que tanto la entrada de datos como la salida de la información es la estándar.

```
input { stdin { } }
output {
  stdout { codec => rubydebug }
}
```

- Ejecutar la configuración básica creada ./bin/logstash -f logstash.conf

```
elk@elk-VirtualBox:~/siem/logstash-6.2.4$ ./bin/logstash -f logstash.conf
Sending Logstash's logs to /home/elk/siem/logstash-6.2.4/logs which is now configured via
log4j2.properties
……
……
……...
[2018-05-05T12:39:22,135][INFO ][logstash.inputs.metrics  ] Monitoring License OK
hola mundo
{
          "message" => "hola mundo",
          "host" => "elk-VirtualBox",
          "@timestamp" => 2018-05-05T10:39:40.851Z,
          "@version" => "1"
}
^C[2018-05-05T12:39:54,636][WARN ][logstash.runner                ] SIGINT received. Shutting
down.
[2018-05-05T12:39:56,593][INFO ][logstash.pipeline          ] Pipeline has terminated
{:pipeline_id=>".monitoring-logstash", :thread=>"#<Thread:0x6acbd555 run>"}
[2018-05-05T12:39:56,757][INFO ][logstash.pipeline          ] Pipeline has terminated
{:pipeline_id=>"main", :thread=>"#<Thread:0x92c2d99 run>"}
elk@elk-VirtualBox:~/siem/logstash-6.2.4$
```

Cofiguración de seguridad: https://www.elastic.co/guide/en/logstash/6.x/ls-security.html

13.4 Instalar Beats

Beats es un agente ligero para la recolección de información, hay multitud de versiones desarrolladas para distintos tipos de plataformas, con distintas funcionalidades, en este TFM vamos a utilizar Filebeat y Metricbeat.

Filebeat
Se ha desarrollado para la obtención de información de ficheros de logs. Los pasos seguidos para su instalación han sido los siguientes:

- Descarga del fichero TAR y descomprimir: https://www.elastic.co/downloads/beats/filebeat

- Configuración del fichero filebeat.yml

```
###################### Filebeat Configuration Example #########################

#========================== Filebeat prospectors
=============================

filebeat.prospectors:
- type: log
  enabled: false
  # Paths that should be crawled and fetched. Glob based paths.
  paths:
          - /var/log/*.log


#============================== Filebeat modules
===============================

filebeat.config.modules:
  # Glob pattern for configuration loading
  path: ${path.config}/modules.d/*.yml

  # Set to true to enable config reloading
  reload.enabled: false

  # Period on which files under path should be checked for changes
  #reload.period: 10s

#================================== Outputs
===================================

# Configure what output to use when sending the data collected by the beat.

#---------------------------- Elasticsearch output ------------------------------
output.elasticsearch:
  # Array of hosts to connect to.
  hosts: ["localhost:9200"]

  # Optional protocol and basic auth credentials.
  #protocol: "https"
  #username: "elastic"
  #password: "changeme"

#----------------------------- Logstash output --------------------------------
#output.logstash:
  # The Logstash hosts
  #hosts: ["localhost:5044"]

  # Optional SSL. By default is off.
  # List of root certificates for HTTPS server verifications
  #ssl.certificate_authorities: ["/etc/pki/root/ca.pem"]
```

```
# Certificate for SSL client authentication
#ssl.certificate: "/etc/pki/client/cert.pem"

# Client Certificate Key
#ssl.key: "/etc/pki/client/cert.key"
```

La parte principal de la configuración es el prospector, que indica el origen de los ficheros a recolectar, a través de la variable path /var/log/*.log, lo que significa que Filebeat recolectará todos los ficheros de la ruta /var/log/ que terminen con la extensión "*./log"*.

Filebeat permite enviar la información directamente a elasticsearch o a logstash, estos parámetros se configuran en el apartado de Output.

En el apartado modules se pueden especificar los patrones de información que contienen los fichero a tratar.

El proceso se inicia con el siguiente comando:

```
sudo ./filebeat -e -c filebeat.yml
```

https://www.elastic.co/guide/en/beats/filebeat/current/filebeat-configuration.html

Metricbeat

Es un agente para la extracción de información del sistema, memoria, procesador, uso del almacenamiento. La opción de instalación elegida a sido la siguiente:

- Descarga y descompresión del fichero TAR : https://www.elastic.co/downloads/beats/metricbeat
- Editar el fichero de configuración *metricbeat.yml*

```
######################## Metricbeat Configuration Example #########################

# This file is an example configuration file highlighting only the most common
# options. The metricbeat.reference.yml file from the same directory contains all the
# supported options with more comments. You can use it as a reference.


#========================  Modules configuration
============================

metricbeat.config.modules:
  # Glob pattern for configuration loading
  path: ${path.config}/modules.d/*.yml   # en esta ruta se ven los activos

  # Set to true to enable config reloading
```

```
  reload.enabled: false

  # Period on which files under path should be checked for changes
  #reload.period: 10s


#==================== Elasticsearch template setting ==========================

setup.template.settings:
  index.number_of_shards: 1
  index.codec: best_compression
  #_source.enabled: false

#============================== Kibana
=====================================

# Starting with Beats version 6.0.0, the dashboards are loaded via the Kibana API.
# This requires a Kibana endpoint configuration.
setup.kibana:

  # Kibana Host
  # Scheme and port can be left out and will be set to the default (http and 5601)
  # In case you specify and additional path, the scheme is required: http://localhost:5601/path
  # IPv6 addresses should always be defined as: https://[2001:db8::1]:5601
  host: "localhost:5601"

#============================== Outputs
===================================

# Configure what output to use when sending the data collected by the beat.

#---------------------------- Elasticsearch output ------------------------------
output.elasticsearch:
  # Array of hosts to connect to.
  hosts: ["localhost:9200"]

  # Optional protocol and basic auth credentials.
  #protocol: "https"
  username: "elastic"
  password: "elastic"

#----------------------------- Logstash output --------------------------------
#output.logstash:
  # The Logstash hosts
  #hosts: ["localhost:5044"]
```

En la parte de Modules se eligen los componentes y el sistema a monitorizar, mientras que en la parte de Outputs se puede elegir enviar los datos a logstash para que sean tratados o almacenados directamente en elasticsearch.

El proceso se inicia con el siguiente comando:

```
sudo ./metricbeat -e -c metricbeat.yml
```

Otros componentes

También es necesario que en el sistema haya otros componentes instalados, estos son:

- Java8 : *apt install openjdk-8-jre-headless*
- cURL : *apt install curl*
- Python 2.7 : *wget --no-check-certificate https://www.python.org/ftp/python/2.7.11/Python-2.7.11.tgz*

14. Anexo II – Alerta primera ejecución de un proceso

Este Watch tiene como objetivo alertar si un proceso se inicia en un servidor por primera vez.

El reloj examina los N minutos previos para los procesos iniciados. Esta lista se usa a su vez para buscar datos anteriores a N minutos, para ver si los procesos se han iniciado históricamente. Cualquier diferencia da como resultado una alerta.

Definición del Watcher

```
{
  "metadata": {
    "window_period": "30s"
  },
  "trigger": {
    "schedule": {
      "interval": "30s"
    }
  },
  "input": {
    "chain": {
      "inputs": [
        {
          "started_processes": {
            "search": {
              "request": {
                "indices": [
                  "log"
                ],
                "body": {
                  "query": {
                    "bool": {
                      "must": [
```

```
                    {
                      "range": {
                        "@timestamp": {
                          "gte": "now-{{ctx.metadata.window_period}}"
                        }
                      }
                    },
                    {
                      "term": {
                        "event_type": {
                          "value": "process_started"
                        }
                      }
                    }
                  ]
                }
              },
              "aggs": {
                "process_hosts": {
                  "terms": {
                    "field": "process_host",
                    "size": 1000
                  }
                }
              },
              "size": 0
            }
          }
        }
      },
      {
        "history_started_processes": {
          "search": {
            "request": {
              "indices": [
                "log"
              ],
              "body": {
                "query": {
                  "bool": {
                    "must": [
                      {
                        "terms": {
                          "process_host": [

"{{#ctx.payload.started_processes.aggregations.process_hosts.buckets}}{{key}}",

"{{/ctx.payload.started_processes.aggregations.process_hosts.buckets}}"
                          ]
```

```
                    }
                  },
                  {
                    "range": {
                      "@timestamp": {
                        "lt": "now-{{ctx.metadata.window_period}}"
                      }
                    }
                  },
                  {
                    "term": {
                      "event_type": {
                        "value": "process_started"
                      }
                    }
                  }
                ]
              }
            },
            "aggs": {
              "process_hosts": {
                "terms": {
                  "field": "process_host",
                  "size": 10
                }
              }
            },
            "size": 0
          }
        }
      }
    }
  ]
 }
},
"condition": {
  "script": {
    "id":"condition"
  }
},
"transform": {
  "script": {
    "id":"transform"
  }
},
"actions": {
  "log": {
    "logging": {
      "text":                                                    "Processes
```

```
started:{{#ctx.payload._value}}{{.}}:{{/ctx.payload._value}}"
      },                                    # se establece la acción a realizar
    "send_email": {
          "email": {
          "profile": "standard",
          "to": [
          "mi_elk_siem@outlook.com"
   }
  }
}
```

Requisitos del Mapping

Esta alerta se puede ejecutar sobre distintos tipos de datos, el requisito es que contengan los tres campos siguientes:

- Marca de tiempo del mensaje de log @timestamp
- Cadena de texto para indicar el host del proceso, es un campo concatenado con el nombre del proceso, y el nombre del host: “process_host”
- Cadena de texto para indicar la actividad que se ha realizado sobre el proceso “event_type”

```
{
  "mappings": {
    "doc": {
      "properties": {
        "@timestamp": {
          "type": "date"
        },
        "process_host": {
          "type": "keyword"
        },
        "event_type": {
          "type": "keyword"
        }
      }
    }
  }
}
```

Requisitos de los datos

El watcher asume que cada documento en elasticsearch representa un proceso en el servidor.
También es un requisito que estén indexados sobre “log” y que el tipo de documento sea “doc”.

Resumen del juego de datos; El juego de datos completo está en la documentación adjunta a la memoria, en la carpeta de test

```
{
        "offset":-600,
        "process_host":"processA_serverA",
        "process":"processA",
        "host":"serverA",
        "event_type":"process_started"
 },
 {
        "offset":-500,
        "process_host":"processA_serverB",
        "process":"processA",
        "host":"serverB",
        "event_type":"process_started"

 },
 {
        "offset":-400,
        "process_host":"processB_serverA",
        "process":"processB",
        "host":"serverA",
        "event_type":"process_started"
 },
 ……..
```

Resultado esperado

Con este juego de datos el resultado esperado es:

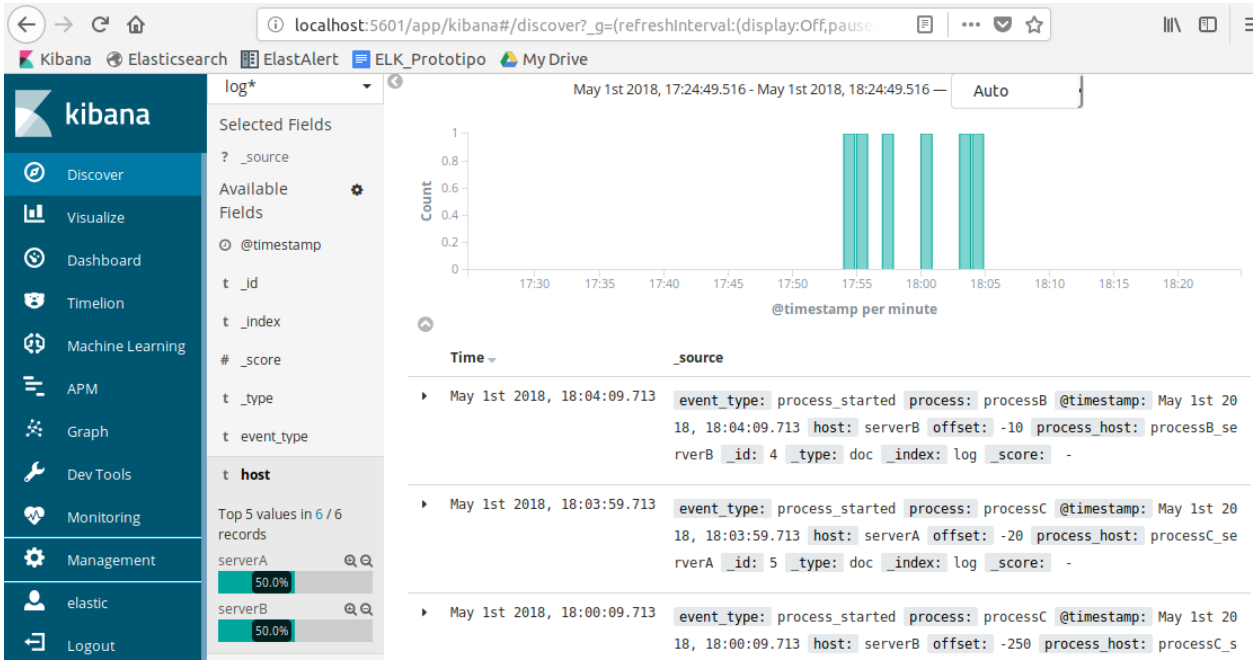
```
"Processes started:processB_serverB:processC_serverA:"
```

Configuración

Los siguientes metadatos del Watch influyen en el comportamiento:

- Window_period - El periodo de N minutos sobre el cual el Watch chequea si hay nuevos procesos iniciados. Este debería ser igual al intervalo programado, por defecto 30 segundos.

Pantallazo de los datos: Index log*



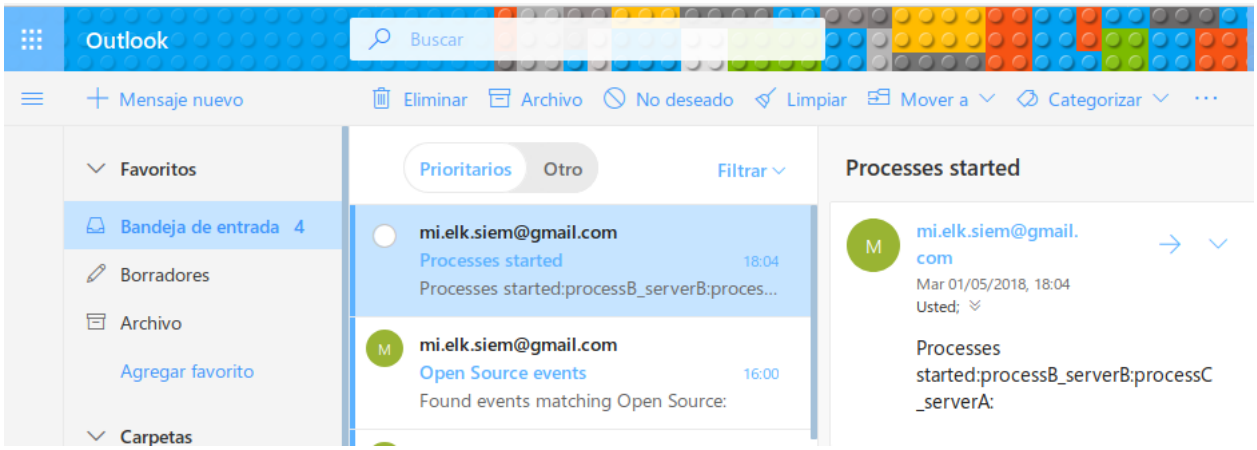
Il·lustració 15 - Datos de procesos cargados

Para la carga de los datos y del Watch dentro del sistema es necesario ejecutar el siguiente script en phyton:

```
elk@elk-VirtualBox:~/siem/examples/Alerting/Sample Watches$ python run_test.py --test_file new_process_started/tests/test1.json --user elastic --password elastic
Expected: Watch Condition: True
Received: Watch Condition: True
Expected: Processes started:processB_serverB:processC_serverA:
Received: Processes started:processB_serverB:processC_serverA:
TEST PASS
elk@elk-VirtualBox:~/siem/examples/Alerting/Sample Watches$
```

Accion desencadenada

Una vez que se han cargado los datos y el Watcher que ejecutará la alerta podemos ver que se ha notificado por correo electrónico el aviso programado:



Il·lustració 16 - Notificación alerta proceso

15. Anexo III – Alerta de escaneo de puertos

El objetivo de este Watcher es detectar y alertar si un servidor establece una gran cantidad de conexiones a un destino, a través de una gran cantidad de puertos.

Una escaneo de puertos se produce cuando se establecen una gran cantidad de conexiones entre dos servidores, en una gran cantidad de puertos distintos. Esto se puede detectar como una gran cantidad de documentos en elasticsearch, con valores de puerto únicos, para los mismos valores de origen-destino.

Esta alerta evita asociar un valor exacto a "alto". En su lugar, pretende basar la interpretación de alto en los datos disponibles y el comportamiento habitual. Además, esta alerta debería poder hacer frente a una gran cantidad de dispositivos> 100k.

Definición del Watcher

El código completo está en la documentación adjunta, aquí solamente se resaltan las partes más importantes

```
{
  "metadata": {
    "time_period": "1m",
    "time_window": "30m",
    "sensitivity": 2.0
  },
  "trigger": {
    "schedule": {
      "interval": "1m"
    }
  },
  "input": {
    "search": {
      "request": {
        "indices": "connection",
        "body": {
          "query": {
            "range": {
              "@timestamp": {
                  "gte": "now-{{ctx.metadata.time_window}}"
                }
            }
          },
          "aggs": {
            "source_dest": {
              "terms": {
```

```
            "field": "source_dest",
            "size": 100,
            "order": {
              "unique_ports": "desc"
            }
          },
          "aggs": {
            "series": {
              "date_histogram": {
                "field": "@timestamp",
                "interval": "{{ctx.metadata.time_period}}"
              },
              "aggs": {
                "num_ports": {
                  "cardinality": {
                    "field": "dest_port"
                  }
                }
              }
            },
            "port_stats": {
              "extended_stats_bucket": {
                "buckets_path": "series>num_ports"
              }
            },
            "median_ports": {
              "percentiles_bucket": {
                "buckets_path": "series>num_ports",
                "percents": [
                  50.0
                ]
              }
            },
            "unique_ports": {
              "cardinality": {
                "field": "dest_port"
              }
            }
          }
        }
      },
      "size": 0
    }
  }
 }
},
"throttle_period": "1m",
"condition": {
  "script": {
    "id": "condition"
```

```
    }
  },
  "actions": {
         "email_me": {
         "throttle_period_in_millis": 600000,
         "send_email": {
         "email": {
         "profile": "standard",
         "to": [
         "mi_elk_siem@outlook.com"
         ],
         "subject": "Port scan detected",
    "log": {
      "transform": {
        "script": {
          "id":"log_transform"
        }
      },
      "logging": {
        "text": "Port scan detected:{{#ctx.payload._value}}{{.}}:{{/ctx.payload._value}}"
      }
    },
    "index_payload": {
      "transform": {
        "script": {
          "id":"index_transform"
        }
      },
      "index": {
        "index": "connection-scans",
        "doc_type": "scan"
      }
    }
  }
}
```

Requisitos del Mapping

El Watcher proporcionado necesita que al menos los siguientes campos estén dentro del mapping:

- source_dest (cadena no analizada): contiene el origen y el destino de la comunicación como una cadena concatenada, p. testServerA_testServerB. Watch asume que el delimitador es un _ char.
- @timestamp (campo de fecha) - Fecha del mensaje de registro.
- source_dest_port (cadena no analizada): contiene el origen, el destino y el puerto de la comunicación como una cadena concatenada, p.

testServerA_testServerB_5002. Watch asume que el delimitador es un _ char.

- dest_port (integer) - puerto en el que se produjo la comunicación.

```
{
  "mappings": {
          "doc": {
          "properties": {
          "source_dest": {
          "type": "keyword"
          },
          "source_dest_port": {
          "type": "keyword"
          },
          "port": {
          "type": "integer"
          },
          "@timestamp":{
          "type":"date"
        }
      }
    }
  }
}
```

Requisitos de los datos

El Watcher asume que cada documento en elasticsearch representa una comunicación entre dos servidores con al menos el mapeo de datos anterior. Otros requisitos son que el inex se llame “connection” y que el tipo de documento sea “doc”.

Resumen del juego de datos; El juego de datos completo está en la documentación adjunta a la memoria, en la carpeta de test

```
{
          "offset":-1800,
          "source_dest":"hostA_hostB",
          "source_dest_port":"hostA_hostB_5043",
          "dest_port":5043
  },
  {
          "offset":-1800,
          "source_dest":"hostA_hostB",
          "source_dest_port":"hostA_hostB_5053",
          "dest_port":5053
  },
```

```
 {
        "offset":-1740,
        "source_dest":"hostA_hostB",
        "source_dest_port":"hostA_hostB_5043",
        "dest_port":5043
 },.....
…….
……...
```

Resultado esperado

Con este juego de datos el resultado esperado es:

```
"Port scan detected:hostA to hostB:"
```

Configuración

Cada periodo de tiempo, time_period (predeterminado 1 m) el Watcher se ejecuta e identifica las comunicaciones entre dos servidores, que hayan utilizado la mayor cantidad de puertos en la última time_window (30 m por defecto). Esto se logra utilizando términos agregados sobre el campo source_dest ordenando por cardinalidad del dest_port.

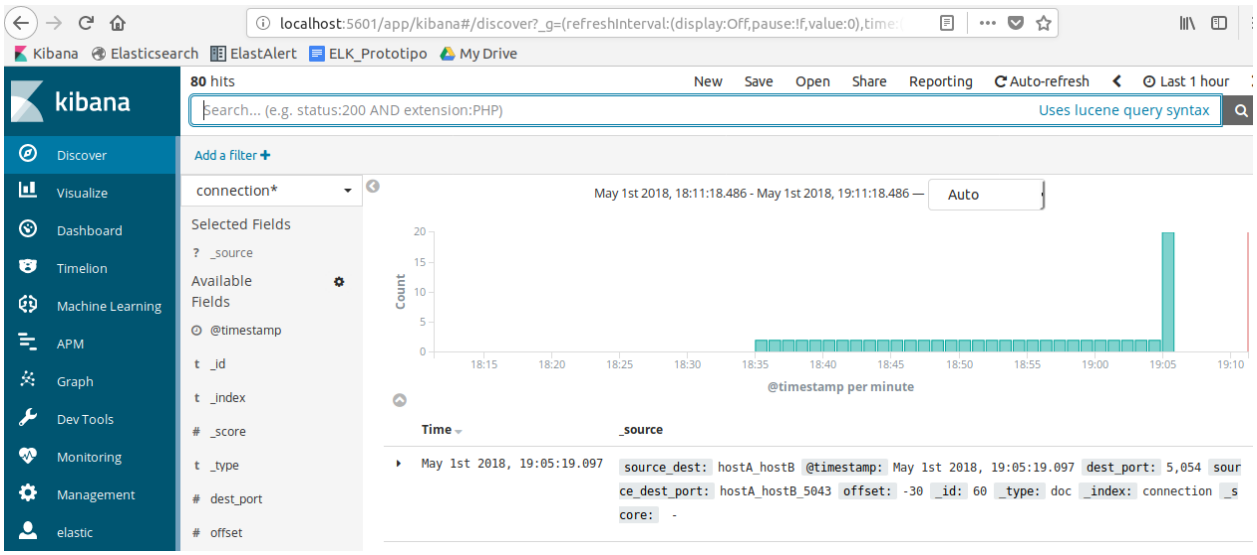
Se considera que un escaneo de puertos ocurre entre dos hosts cuando el número de puertos en el último time_period (es decir, el último segmento del perfil) es 2 veces la desviación estándar por encima de la mediana.

Los siguientes parámetros de metadatos de observación influyen en el comportamiento:

- time_window - El período N (minutos) durante el cual la mediana y el estándar dev. se calcula para cada par source_dest. Predeterminado a 30m.
- time_period: el período X (minutos) o el tamaño de cada segmento. Esto define el período más pequeño en el que se puede detectar un escaneo de puerto. Predeterminado a 1m.

La cantidad de segmentos utilizados para calcular el promedio afectará significativamente el rendimiento, es decir, el tamaño de la window_size/time_period.

Pantallazo Index connection*



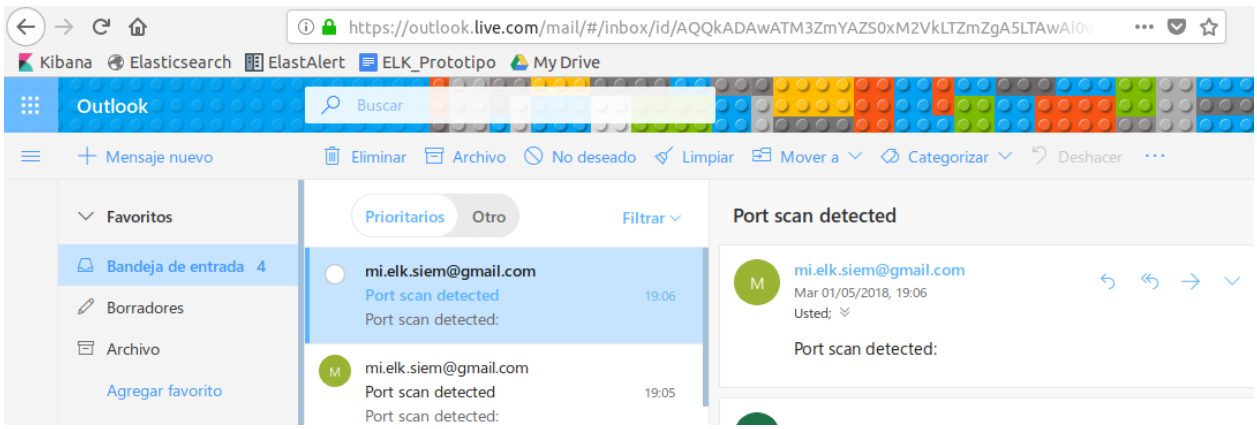
Il·lustració 17 - Datos Alerta escaneo de puertos

La ejecución del siguiente script ejecuta la carga del mapping, los datos y el Watcher en el sistema:

```
elk@elk-VirtualBox:~/siem/examples/Alerting/Sample Watches$ python run_test.py --test_file port_scan/tests/test1.json --user elastic --password elastic
Unable to delete current dataset
Expected: Watch Condition: True
Received: Watch Condition: True
Expected: Port scan detected:hostA to hostB:
Received: Port scan detected:hostA to hostB:
TEST PASS
elk@elk-VirtualBox:~/siem/examples/Alerting/Sample Watches$
```

Accion desencadenada

Como resultado, se ha recibido el siguiente correo:



Il·lustració 18 - Notificación Alerta escaneo de puertos

16. Anexo IV – Alerta de actividad de cuenta sospechosa

Este Watcher trata de detectar y alertar si un usuario es creado en Active Directory / LDAP y posteriormente eliminado dentro de N minutos siguientes.

Definición del Watcher

El código completo está en la documentación adjunta, aquí solamente se resaltan las partes más importantes

```
{
  "metadata": {
    "window_period": "5m"
  },
  "trigger": {
    "schedule": {
      "interval": "1m"
    }
  },
  "input": {
    "search": {
      "request": {
        "indices": "logins",
        "types": "doc",
        "body": {
          "query": {
            "bool": {
              "must": [
                {
                  "range": {
                    "@timestamp": {
                      "gte": "now-{{ctx.metadata.window_period}}"
                    }
                  }
                }
              ],
              "should": [
                {
                  "terms": {
                    "event_type": [
                      "add",
                      "remove"
                    ]
                  }
```

```
            }
          ]
        }
      },
      "aggs": {
        "event_types": {
          "filters": {
            "filters": {
              "add": {
                "term": {
                  "event_type": {
                    "value": "add"
                  }
                }
              },
              "remove": {
                "term": {
                  "event_type": {
                    "value": "remove"
                  }
                }
              }
            }
          },
          "aggs": {
            "users": {
              "terms": {
                "field": "user",
                "size": 1000
              }
            }
          }
        }
      },
      "size": 0
    }
  }
},
"condition": {
  "script": {
    "id": "condition"
  }
},
"throttle_period": "5m",
"actions": {
  "log": {
    "transform": {
      "script": {
        "id":"transform"
```

```
      }
    },
    "logging": {
      "text": "Users added and removed in the last {{ctx.metadata.window_period}}:{{#ctx.payload._value}}{{.}}:{{/ctx.payload._value}}"
    },
        "send_email": {
        "email": {
        "profile": "standard",
        "to": [
        "mi_elk_siem@outlook.com"
        ],
        "subject": "Users added and removed"
   }
  }
}
```

Requisitos del Mapping

El Watcher requiere que los datos de entrada al menos tengan los siguientes campos:

- event_type (non-analyzed string) -contiene la entrada del AD, el tipo de evento tiene que tener los valores "add" o "remove".
- @timestamp (date field) - Fecha del mensaje de log.
- user (non-analyzed string) - El id del usuario cuya operación se ha realizado.

Mapping

```
{
  "mappings": {
        "doc": {
        "properties": {
        "event_type": {
        "type": "keyword"
        },
        "user": {
        "type": "keyword"
        },
        "@timestamp":{
        "type":"date"
        }
      }
   }
  }
}
```

Requisitos de los datos

El Watcher asume que cada documento en Elasticsearch representa una registro del log, de la actividad en Active Directory o LDAP, independientemente de la forma de recolección de los datos.
Todos los elementos son indexados en “logins” y el tipo de documento es “doc”.
Se supone un máximo de 1000 usuarios monitorizados concurrentemente.

Resumen del juego de datos; El juego de datos completo está en la documentación adjunta a la memoria, en la carpeta de test.

```
  {
        "offset":-200,
        "user":"testUserB",
        "event_type":"add"
 },
 {
        "offset":-100,
        "user":"testUserC",
        "event_type":"add"
 },
 {
        "offset":-45,
        "user":"testUserA",
        "event_type":"remove"
 },
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
```

Resultado esperado

Con este juego de datos el resultado esperado es:

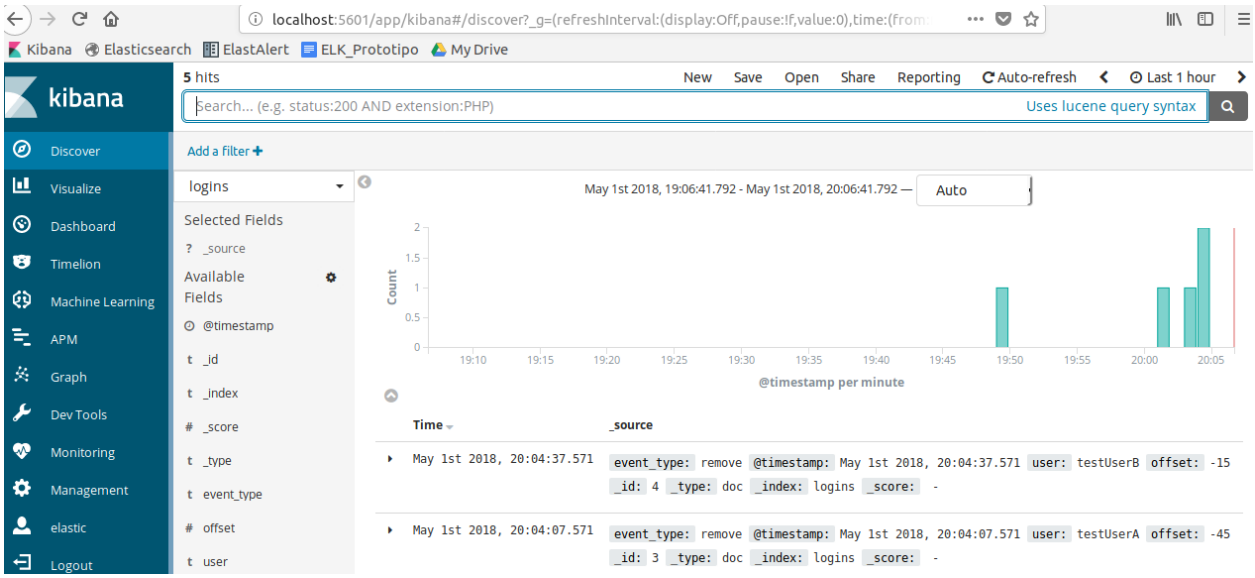
```
"expected_response":"Users added and removed in the last 5m:testUserB:"
```

Configuración

Los metadatos del Watcher parametrizables pueden influir en el comportamiento:

- window_period - El periodo N (mins) dentro del cual una cuenta de usuairo tiene que ser creada y posteriormente borrada. Por defecto se establecen 5 minutos.

Pantallazo de los datos Index logins



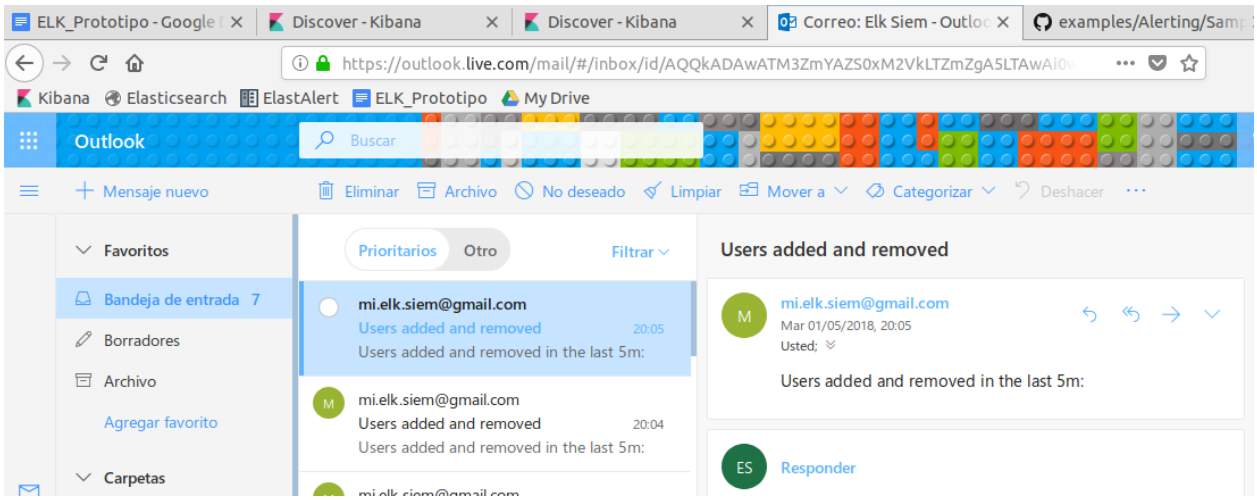
Il·lustració 19 - Datos de creación de usuarios

La ejecución del siguiente script carga los datos, el mapping y el Watcher en el sistema:

```
elk@elk-VirtualBox:~/siem/examples/Alerting/Sample Watches$ python run_test.py --test_file unexpected_account_activity/tests/test1.json --user elastic --password elastic
Unable to delete current dataset
Expected: Watch Condition: True
Received: Watch Condition: True
Expected: Users added and removed in the last 5m:testUserB:
Received: Users added and removed in the last 5m:testUserB:
TEST PASS
elk@elk-VirtualBox:~/siem/examples/Alerting/Sample Watches$
```

Accion desencadenada

Correo recibido:



Il·lustració 20 - Notificación alerta de creación y borrado de usuarios

Referencia:
https://github.com/elastic/examples/tree/master/Alerting/Sample%20Watches/unexpected_account_activity

15. Anexo V – Alerta analizando SSH

En este ejemplo se van a utilizar datos en el formato CEF, para analizar los eventos de log en los inicios de sesión de un servicio de SSH.
La finalidad es detectar los logins correctos desde una dirección externa, y también detectar los ataques de fuerza bruta que puede recibir el sistema, definiendo esta situación como una secuencia de N intentos fallidos de login, seguidos de un login exitoso.

Definición del Watcher 1

Este Watcher va a detectar los logins en el sistema desde una IP remota.
El código completo está en la documentación adjunta, aquí solamente se resaltan las partes más importantes

```
{
  "trigger": {
    "schedule": {
      "interval": "10s"
    }
  },
  "input": {
    "search": {
      "request": {
        "indices": [
          "<cef-ssh-{now/d}>"
        ],
        "types": [
          "syslog"
        ],
        "body": {
          "query": {
            "bool": {
              "must_not": [
                {
                  "term": {
                    "sourceAddress": {
                      "value": "192.168.0.0/16"
                    }
                  }
                }
              ],
              "filter": [
                {
                  "term": {
                    "categoryBehaviour": "cowrie.login.success"
                  }
                },
```

```
              {
                "range": {
                  "@timestamp": {
                    "gte": "{{ctx.trigger.scheduled_time}}||-10s"
                  }
                }
              }
            ]
          }
        },
        "size": 100,
        "_source": [
          "destinationHostName",
          "sourceAddress",
          "@timestamp"
        ]
      }
    }
  },
  "condition": {
    "compare": {
      "ctx.payload.hits.total": {
        "gt": 0
      }
    }
  },
  "transform": {
    "script": "return ctx.payload.hits.hits.stream().map(d ->
['@timestamp':d._source['@timestamp'],'sourceAddress':d._source.sourceAddre
ss,'destinationHostName':d._source.destinationHostName]).collect(Collectors.to
List());"
  },
  "actions": {
    "index_payload": {
      "transform": {
        "script": "return ['_doc':ctx.payload._value];"
      },
      "index": {
        "index": "cef-ssh-watch-results",
          "doc_type": "external_login"
      }
    }
  }
}
```

Requisitos del Mapping

Este mapping utiliza una plantilla con el formato CEF para trabajar con eventos, el documento completo se encuentra junto con la documentación del TFM (cef_template.json)

```
{
   "order": 100,
   "template": "cef-*",
   "settings": {
     "index": {
       "number_of_shards": "3",
       "number_of_replicas": "0"
     }
   },
   "mappings": {
     "_default_": {
       "dynamic": true,
       "dynamic_templates": [
         {
           "string_fields": {
             "mapping": {
               "type": "keyword"
             },
             "match_mapping_type": "string",
             "match": "*"
           }
         }
       ],
       "_all": {
         "enabled": true
       },
       "properties": {
         "destinationPort": {
           "type": "integer"
         },
         "flexDate1": {
           "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
           "type": "date"
         },
         "sourcePort": {
           "type": "integer"
         },
         "baseEventCount": {
           "type": "integer"
         },
         "destinationAddress": {
           "type": "ip"
         },
         "destinationProcessId": {
           "type": "integer"
         },
         "oldFileSize": {
```

```
      "type": "integer"
    },
    "destination": {
      "dynamic": false,
      "type": "object",
      "properties": {
        "city_name": {
          "type": "keyword"
        },
        "country_name": {
          "type": "keyword"
        },
        "location": {
          "type": "geo_point"
        },
        "region_name": {
          "type": "keyword"
        }
      }
    },
    "source": {
      "dynamic": false,
      "type": "object",
      "properties": {
        "city_name": {
          "type": "keyword"
        },
        "country_name": {
          "type": "keyword"
        },
        "location": {
          "type": "geo_point"
        },
        "region_name": {
          "type": "keyword"
        }
      }
    },
    "deviceReceiptTime": {
      "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
      "type": "date"
    },
    "destinationTranslatedPort": {
      "type": "integer"
    },
    "deviceTranslatedAddress": {
      "type": "ip"
    },
    "deviceAddress": {
      "type": "ip"
    },
    "agentReceiptTime": {
      "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
      "type": "date"
```

```
},
"startTime": {
  "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
  "type": "date"
},
"sourceProcessId": {
  "type": "integer"
},
"bytesIn": {
  "type": "integer"
},
"bytesOut": {
  "type": "integer"
},
"severity": {
  "omit_norms": true,
  "type": "string"
},
"deviceProcessId": {
  "type": "integer"
},
"agentAddress": {
  "type": "ip"
},
"sourceAddress": {
  "type": "ip"
},
"sourceTranslatedPort": {
  "type": "integer"
},
"deviceCustomDate2": {
  "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
  "type": "date"
},
"deviceCustomDate1": {
  "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
  "type": "date"
},
"flexNumber1": {
  "type": "long"
},
"deviceCustomFloatingPoint1": {
  "type": "float"
},
"oldFileModificationTime": {
  "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
  "type": "date"
},
"deviceCustomFloatingPoint2": {
  "type": "float"
},
"oldFileCreateTime": {
  "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
  "type": "date"
```

```
},
"deviceCustomFloatingPoint3": {
  "type": "float"
},
"sourceTranslatedAddress": {
  "type": "ip"
},
"deviceCustomFloatingPoint4": {
  "type": "float"
},
"flexNumber2": {
  "type": "long"
},
"fileCreateTime": {
  "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
  "type": "date"
},
"fileModificationTime": {
  "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
  "type": "date"
},
"fileSize": {
  "type": "integer"
},
"destinationTranslatedAddress": {
  "type": "ip"
},
"endTime": {
  "format": "epoch_millis||epoch_second||date_time||MMM dd yyyy HH:mm:ss",
  "type": "date"
},
"deviceCustomNumber1": {
  "type": "long"
},
"deviceDirection": {
  "type": "integer"
},
"device": {
  "dynamic": false,
  "type": "object",
  "properties": {
    "city_name": {
      "type": "keyword"
    },
    "country_name": {
      "type": "keyword"
    },
    "location": {
      "type": "geo_point"
    },
    "region_name": {
      "type": "keyword"
    }
  }
```

```
        },
        "deviceCustomNumber3": {
          "type": "long"
        },
        "deviceCustomNumber2": {
          "type": "long"
        },
        "categoryOutcome": {
          "type": "keyword"
        },
        "destinationHostName": {
          "type": "keyword"
        },
        "destinationAddress": {
          "type":"ip"
        }
      }
    }
  },
  "aliases": {}
}
```


Requisitos de los datos

Resumen del juego de datos; El juego de datos completo está en la documentación adjunta a la memoria, en la carpeta con el resto de ficheros del ejemplo.


```
CEF:0|Unix|Unix|5.0|cowrie.session.connect|New connection: 192.168.1.105:60740 (192.168.1.105:2222) [session: 6e99ac86]|Unknown|externalId=1 startTime=Nov 15 2016 19:18:21 destinationHostName=elastic_honeypot destinationAddress=192.168.20.2 deviceReceiptTime=Nov 15 2016 19:18:21 deviceTimeZone=Z transportProtocol=TCP applicationProtocol=SSHv2 destinationServiceName=sshd devicePayloadId=1 message=New connection: 192.168.1.105:60740 (192.168.1.105:2222) [session: 6e99ac86] destinationAddress=192.168.1.105 destinationTranslatedAddress=192.168.1.105 deviceTranslatedAddress=192.168.1.105 deviceAddress=192.168.1.105 destinationTranslatedPort=2222 destinationPort=2222 categoryOutcome=None categoryBehaviour=cowrie.session.connect sourceTranslatedAddress=192.168.1.105 sourceAddress=192.168.1.105 sourceTranslatedPort=60740 sourcePort=60740 deviceDirection=1 cs1=0 cs1Label=isError cs2=cowrie.ssh.factory.CowrieSSHFactory cs2Label=system cs4=6e99ac86 cs4Label=session
CEF:0|Unix|Unix|5.0|cowrie.client.version|Remote SSH version: SSH-2.0-OpenSSH_7.2p2 Ubuntu-4ubuntu2.1|Unknown|externalId=2 startTime=Nov 15 2016 19:18:21
```


Resultado esperado

Con los datos de ejemplo se espera detectar tres ataques de fuerza bruta contra el sistema monitorizado.

Configuración

Primero se necesita parametrizar el fichero de configuración de Logstash para que pueda utilizar la plantilla CEF, en la manipulación de los datos.

Segundo se necesita realizar el volcado de la información en el sistema, dada la cantidad de registros (114.151 documentos) la carga puede demorar varios minutos. La orden puede darse desde un terminal:

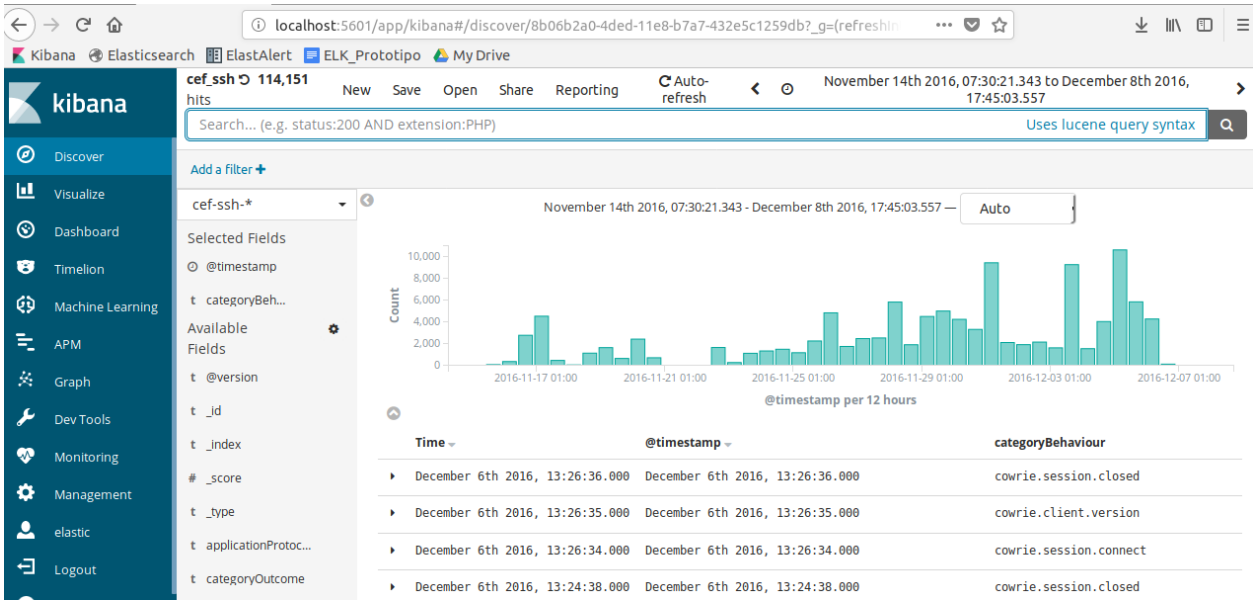
```
elk@elk-VirtualBox:~/siem/examples/Security Analytics/ssh_analysis$ cat ssh.cef | nc localhost 5000
```

Se puede comprobar que los datos se han cargado correctamente a través de los comandos cURL;

```
elk@elk-VirtualBox:~$ curl http://localhost:9200/cef-ssh-*/_count -u elastic:elastic

{"count":114151,"_shards":{"total":105,"successful":105,"skipped":0,"failed":0}}elk@elk-VirtualBox:~$
```

Pantallazo de los datos Index cef*



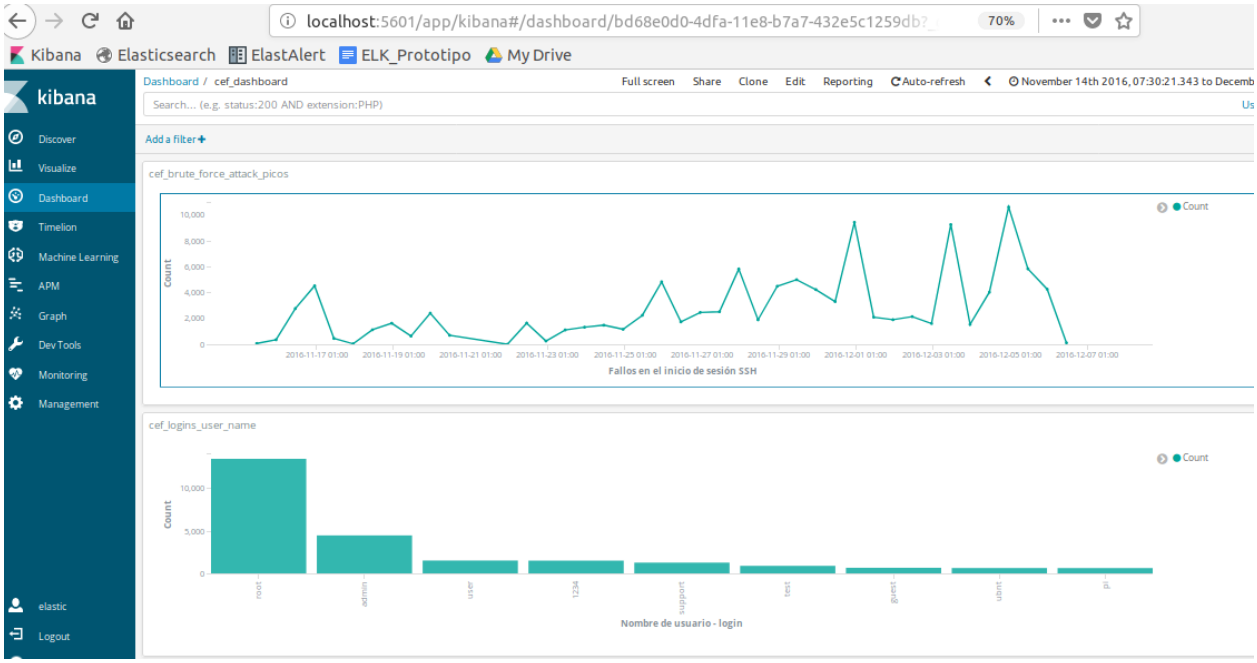
Il·lustració 21 - Datos CEF actividad SSH

El Watch tiene que ser ejecutado sobre el dataset al completo, se realizará desde la consola de comandos:

```
elk@elk-VirtualBox:~/siem/examples/Security Analytics/ssh_analysis$ ./run_watch.sh
successful_login_external.inline elastic elastic
Loading successful_login_external.inline watch
Loading successful_login_external.inline watch...OK
elk@elk-VirtualBox:~/siem/examples/Security Analytics/ssh_analysis$ ./run_watch.sh
brute_force_login.inline  elastic elastic
Loading brute_force_login.inline watch
Loading brute_force_login.inline watch...OK
elk@elk-VirtualBox:~/siem/examples/Security Analytics/ssh_analysis$
```

Accion desencadenada

Los resultados de la ejecución podemos verlos en Kibana, en el dashboard "cdf_dashboard"



Il·lustració 22 - Gráficas del analisis de los datos CEF SSH

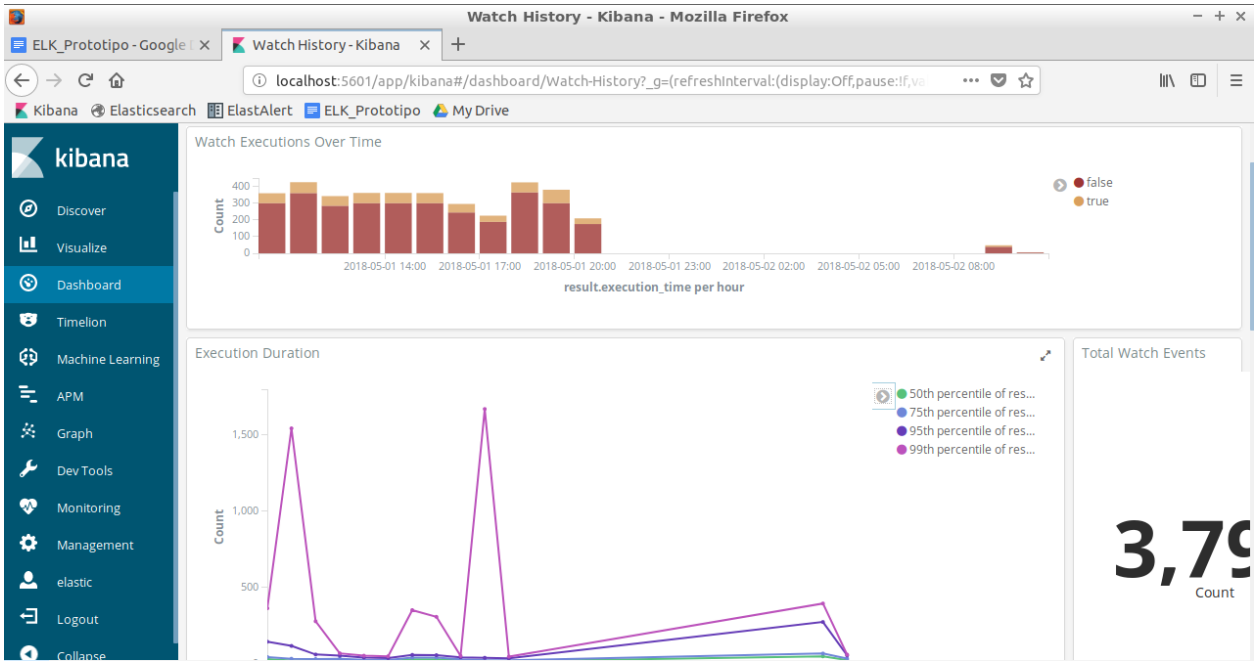
La gráfica superior muestra 3 picos a la derecha que se corresponden con los intentos de ataque por fuerza bruta al sistema.

La gráfica inferior representa los nombres de usuario más utilizados para realizar los ataques, en primer lugar "root", en segundo lugar "admin", y en tercer lugar "user".

16. Anexo VI – Cuadro de mando

Ver los resultados de forma gráfica ayuda mucho a comprender el estado del sistema que se está monitorizando, con esta finalidad se han agrupado los eventos que han lanzado los distintos Watcher.

El cuadro de mando se ha parametrizado utilizando uno de los ejemplos que se proporcionan con el plug-in X-pack, en el que se permite ver la historia de los Watcher que han estado en ejecución en el sistema.



Il·lustració 23 - Cuadro de mando

La primera gráfica representa los watcher que se han ejecutado en un periodo de tiempo, diferenciando en rojo aquellos de resultado negativo y en naranja aquellos de resultado positivo.

La segunda gráfica representa el tiempo de ejecución de los Watcher, nos sirve para monitorizar el rendimiento del sistema y su carga de trabajo.

Los resultados pueden ser llamativos, debido a que es un entorno de laboratorio en el que se han buscado ejemplos para que se lancen las alarmas de una forma rápida.

Referencia:
https://github.com/elastic/examples/tree/master/Alerting/watcher_dashboard

[i] Licencia Apache - https://en.wikipedia.org/wiki/Apache_License
[ii] Términos clave - Applied Network Security Monitoring: Collection, Detection & Analysis (ISBN - 978-0124172081)
[iii] SANS Institute InfoSec Reading Room: Implementing the 20 Critical Controls with Security Information and Event Management (SIEM) Systems https://www.sans.org/reading-

room/whitepapers/analyst/implementing-20-critical-controls-security-information-event-management-siem-systems-34965

[iv] Logs, trees, forest - Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management ISBN 978-1597496353

[v] Categorías de los mensajes de Log - Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management ISBN 978-1597496353

[vi] How is Log Data Transmitted and Collected? - Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management - ISBN 978-1597496353

[vii] Basic tools for log analysis - Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management - ISBN 978-1597496353

[viii] Más información de la herramienta OSSEC en la web www.ossec.net

[ix] Más información de la herramienta OSSIM en la web http://www.alienvault.com/

[x] Web de Splunk www.splunk.com

[xi] NetIQ Sentinel https://www.netiq.com/products/sentinel

[xii] IBM q1Labs QRadar http://q1labs.com/products.aspx

[xiii] Loggly http://loggly.com/

[xiv] SANS Institute InfoSec Reading Room: Implementing the 20 Critical Controls with Security Information and Event Management (SIEM) Systems https://www.sans.org/reading-room/whitepapers/analyst/implementing-20-critical-controls-security-information-event-management-siem-systems-34965

[xv] Center for Internet Security: https://www.cisecurity.org/controls/

[xvi] Tools for Log Analysis - Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management ISBN 978-1597496353

[xvii] Elasticsearch - https://en.wikipedia.org/wiki/Elasticsearch

[xviii] Logstash - https://www.elastic.co/products/logstash

[xix] Kibana - https://en.wikipedia.org/wiki/Kibana

[xx] Plugin Watcher - https://www.elastic.co/guide/en/watcher/current/introduction.html