Análisis de Amenazas de Seguridad Utilizando el Stack de Elasticsearch

ANUAR LEZAMA BARQUET
Especialidad de Big Data

En México durante el año 2018 se registraron diversos ciberataques a numerosas instituciones de sector financiero. Entre estos ciberataques destacan el intento que se realizó al Banco Nacional de Comercio (Bancomext) y el robo de alrededor de millones de pesos a cinco instituciones financieras a través del Sistema de Pagos Electrónicos (SPEI). La sofisticación de los ataques, la mejor preparación de los atacantes y el tiempo que dedican en la planeación de estos plantea la necesidad de que las organizaciones estén más conscientes y preparadas para poder detectar amenazas de seguridad lo antes posible.

Los Sistemas de Gestión de Eventos e Información de Seguridad (Security Information and Event Management - SIEM) son herramientas que ayudan en implementar la seguridad informática en una organización mediante la recolección de información relacionada con la seguridad de la infraestructura de TI.

El núcleo de los sistemas SIEM se basa en un punto central de agregación de logs junto con un sistema para el análisis de estos. Los logs pueden proceder de aplicaciones, firewalls, bases de datos y sistemas de red, esta información en crudo permite conocer el estado del ambiente de TI y ayudar en la identificación de anomalías que puedan indicar potenciales brechas de seguridad en la organización.

Mediante el análisis de los logs, los sistemas SIEM permiten, por ejemplo, conocer cuál es la dirección IP de donde está procediendo un ataque, el número de equipos que posiblemente fueron comprometidos, etc.

Los sistemas SIEM deben poseer diversas características las cuales son:

- **Agregación y colección de información**. Los sistemas SIEM deben de tener la capacidad de ser el punto central de agregación de logs de todos los sistemas que componen la infraestructura de TI de manera que puedan consolidar la información con el fin de evitar que determinados eventos críticos no sean detectados. Adicionalmente, debe de ser capaces de soportar una variedad de protocolos de comunicación para poder conectar todos los sistemas de la organización.
- **Retención**. Los sistemas SIEM deben de ser capaces de retener la información para poder realizar la correlación de la información con información histórica.
- **Correlación**. El sistema SIEM debe de ser capaz de correlacionar eventos de diferentes fuentes, y empaquetarlos permitiendo ser visualizados de manera conjunta.

- **Alertamiento**. Un sistema SIEM debe ser capaz de alertar cuando la correlación de eventos indica una posible amenaza o comportamiento fuera de lo normal. Idealmente, el sistema también debe de ser capaz de utilizar la información histórica para entrenar sistemas de aprendizaje de máquina que ayuden en la detección de anomalías y que automaticen el trabajo de detección de estas.
- **Dashboards**. Las herramientas de visualización de los sistemas SIEM basadas en dashboards permiten a través de indicadores y gráficas asistir a los analistas en identificar patrones y actividades que no entran dentro del estándar. Estos dashboards se alimentan con información en tiempo real lo cual permite un monitoreo continuo.
- **Análisis forense**. Los sistemas SIEM deben de tener la habilidad de realizar búsquedas de diferentes periodos de tiempo bajo un determinado criterio.

Si bien comercialmente existen diversos sistemas SIEM, algunos Open Source y que incluso utilizan el ELK como su núcleo, en esta sección se explica la razón por la que ELK puede ser utilizado como una solución SIEM. El stack de Elasticsearch, conocido como ELK por Elasticsearch, Logstash y Kibana, más Beats y el X-Pack permiten, en su conjunto cumplir con las funciones que requiere un sistema SIEM.

Agregación y colección de información. La combinación de Logstash más Beats permite obtener información de diferentes fuentes mediante el uso de pipelines. Beats posee componentes para la captura de archivos de logs, capturar el tráfico de red en diferentes protocolos como HTTP, o métricas del estado de los servidores esta información puede ser preprocesada por el Beat y posteriormente ser enviada hacia Logstash para realizar parseo y transformaciones o directamente ser enviada a Elasticsearch. La ventaja de uso de Beats es que es de simple instalación y configuración además de generar un mínimo impacto en el sistema.

Retención. Elasticsearch permite almacenar e indexar la información. Elasticsearch se ha convertido en una de las bases de datos Open Source más populares lo cual se debe a su facilidad de uso, de instalación, escalabilidad, tolerancia a fallos y un gran apoyo para su soporte por parte de la comunidad.

Correlación. Como tal ELK no posee un módulo de correlación. La correlación de diversos eventos en ELK debe de ser realizado de manera manual mediante la definición de consultas por medio de Kibana. Las consultas de Kibana permiten por ejemplo, identificar si existen una determinada cantidad de peticiones en un intervalo de tiempo. Otro ejemplo es el poder crear una regla que correlacione el número de log-ins fallidos con la creación de cuentas con privilegios.

Alertamiento. El X-Pack posee un módulo de alertas el cual puede realizar notificaciones por correo, chat, slack, entre otros y posee un webhooc que puede ser integrado con la

infraestructura de monitoreo existente. Adicionalmente, el sistema de alertas puede no sólo alimentarse de reglas estáticas, sino que también puede ser conectando al X-Pack de machine learning.

Dashboards. Kibana permite la creación de dashboards con la información que se encuentra en Elasticsearch. Kibana posee las gráficas clásicas de visualización como histogramas, graficas de líneas, pastel, aparte de poseer Vega grammar para poder realizar visualizaciones a la medida.

Análisis forense. La capacidad de escalabilidad de Elasticsearch tanto de manera horizontal como vertical permite almacenar información histórica suficiente para la detección de ataques que se desarrollan durante un periodo de tiempo extenso. Adicionalmente, la indexación permite que la información histórica se puede consultar de manera rápida lo cual permite encontrar patrones en poco tiempo.

ELK sirve como parte de la arquitectura de soluciones SIEM como es OSSEC Wazuah, SIEMoster y Apache Metron por lo que su uso para la construcción de un SIEM está comprobado. Utilizar Elasticsearch y Kibana para realizar la visualización y realizar el análisis de la información en otra plataforma.

En Praxis contamos con la especialidad de Big Data que cuenta con personal que conocen y son capaces de utilizar el stack de ELK para la construcción de sistemas de seguridad SIEM, búsqueda de sitios web, búsquedas empresariales y herramientas de análisis de información.

Bibliografía

Berman, D. (2018, June 20). Using the ELK Stack for SIEM. Retrieved March 29, 2019, from https://logz.io/blog/elk-siem/

Deshpande, A., & Kumar, M. (2018). *Artificial Intelligence for big data: Complete guide to automating big data solutions using artificial intelligence techniques*. Birmingham, UK: Packt Publishing.

Johansen, G. (2017). *Digital forensics and incident response: A practical guide to deploying digital forensic techniques in response to cyber security incidents*. Birmingham, UK: Packt Publishing.

Riquelme, R. (2018, December 31). No pinta un 2019 sólido para la ciberseguridad en México. Retrieved from http://www.eleconomista.com.mx/tecnologia/No-pinta-un-2019-solido-para-la-ciberseguridad-en-Mexico-20181229-0003.html.

Paquette, M. (2018, September 05). Using Machine Learning and Elasticsearch for Security Analytics: A Deep Dive. Retrieved from http://www.elastic.co/blog/using-machine-learning-and-elasticsearch-for-security-analytics-deep-dive