INNOVA Research Journal, ISSN 2477-9024
(Septiembre-Diciembre 2021). Vol. 6, No.3.2 pp. 141-167
DOI: https://doi.org/10.33890/innova.v6.n3.2.2021.1860
URL: http://revistas.uide.edu.ec/index.php/innova/index
Correo: innova@uide.edu.ec

Identificación de amenazas informáticas aplicando arquitecturas de Big Data

Identification of IT threats by applying Big Data architectures

Fabián Balseca-Chávez
Universidad Ecotec, Guayaquil, Ecuador
fbalseca@mgs.ecotec.edu.ec; fabianbch@yahoo.com
https://orcid.org/0000-0002-7994-3890

Alejandra Mercedes Colina-Vargas
Universidad Ecotec, Guayaquil, Ecuador
acolina@ecotec.edu.ec
https://orcid.org/0000-0003-1514-8852

Marcos Antonio Espinoza-Mina
Universidad Ecotec, Guayaquil, Ecuador
mespinoza@ecotec.edu.ec
https://orcid.org/0000-0003-1530-7243

Recepción: 26/07/2021 | Aceptación: 09/11/2021 | Publicación: 25/11/2021

Cómo citar (APA, séptima edición):
Balseca-Chávez, F., Colina-Vargas, A. M., y Espinoza-Mina, M.A. (2021). Identificación de amenazas informáticas aplicando arquitecturas de Big Data. *INNOVA Research Journal*, *6*(3.2), 141-167. https://doi.org/10.33890/innova.v6.n3.2.2021.1860

Resumen
El uso masivo de las Tecnologías de la Información y Comunicaciones ha ocasionado la interdependencia de la sociedad respecto de estas; sumado a la ausencia de controles eficientes y efectivos a nivel general, incrementan la exposición a los ataques o amenazas informáticas, a las vulnerabilidades en los activos de información de las organizaciones. En este contexto, el presente artículo propone una arquitectura de análisis de datos a través de herramientas de Big Data mediante la utilización de eventos o registros de seguridad, que permitan mejorar la identificación, integración y correlación de eventos. La metodología de la investigación soportada se caracterizó por ser exploratoria y descriptiva. Para el desarrollo de la solución propuesta se empleó las fases del procesamiento de Big Data propuesta por Labrinidis y Jagadish, que permita la identificación de amenazas de informáticas. La arquitectura tecnológica diseñada se basó en la integración de Elastic Stack y sus componentes principales (Elasticsearch, Logstash, Kibana), y tecnologías como

Filebeat y Wazuh Security Detection (NIPS/HIDS), gestionando la seguridad en activos de información como equipos de comunicaciones, servidores de datos y aplicaciones, motores de bases de datos, y terminales de usuario final. Su implementación permitiría la supervisión en tiempo real e histórica de una respuesta ágil y efectiva de alertas de seguridad e informes de estado ante incidentes.
Palabras claves: seguridad de los datos; tecnología de la información; base de datos; análisis de datos.

Abstract
The massive use of Information and Communication Technologies has caused the interdependence of society with respect to them; added to the absence of efficient and effective controls at a general level, they increase the exposure to attacks or computer threats, to vulnerabilities in the information assets of the organizations. In this context, this article proposes a data analysis architecture through Big Data tools using events or security logs, which allow to improve the identification, integration and correlation of events. The methodology of the supported research was characterized as exploratory and descriptive. For the development of the proposed solution, the phases of Big Data processing proposed by Labrinidis & Jagadish were used, allowing the identification of computer threats. The technological architecture designed was based on the integration of Elastic Stack and its main components (Elasticsearch, Logstash, Kibana), and technologies such as Filebeat and Wazuh Security Detection (NIPS / HIDS), managing security in information assets such as communications equipment, data and application servers, database engines, and end-user terminals. Its implementation would allow real-time and historical monitoring of an agile and effective response to security alerts and incident status reports.
Keywords: data security; information technology; database; data analysis.

Introducción

En la última década, las organizaciones y empresas se han visto inmersas en sofisticados y diversos ataques a los sistemas informáticos, comprometiendo sus datos y activos informáticos, obligándolas a centrar su atención en la gestión y seguridad de la información. El uso masivo de las Tecnologías de la Información y Comunicaciones (TIC) ha causado la dependencia de la sociedad respecto de estas, las empresas no escapan a esta realidad.

En particular, hoy el uso de esa tecnología, incluida la Internet, facilita la vida de las personas, sin embargo, lamentablemente éstas, no están muy familiarizadas con la seguridad; aumentando con ello la posibilidad de ataques (Subburaj et al., 2018). Complejiza este escenario, los problemas para los sistemas de seguridad debido a la ausencia de controles eficientes y efectivos en el uso de las TIC, y a la carencia de normativas legales que regulen la utilización del ciberespacio, lo cual conlleva a que continuamente incrementen las vulnerabilidades en los activos de información de las organizaciones.

Una muestra de lo expuesto, son los cambios radicales y acelerados que se han dado en las organizaciones a raíz de la pandemia provocada por el Coronavirus al inicio del año 2020, obligándolas a dar respuesta a nuevos paradigmas en la forma de interactuar; se cambió la naturaleza del trabajo diario a otra modalidad, el teletrabajo. Modificando con ello, en la mayoría

de las empresas, el ambiente donde se dan las amenazas y vulnerabilidades en el ámbito tecnológico, así como también las responsabilidades del equipo de seguridad.

Constituye hoy, la seguridad de la información uno de los principales motivos para que las empresas en América Latina y en todo el mundo se preparen a sortear entornos complejos, trayendo consigo nuevos retos y desafíos que superar, dado que se mantienen en incremento, incluso después de la pandemia, los incidentes de seguridad a nivel de códigos maliciosos, robo de información y accesos indebidos a los sistemas siendo de interés y preocupación por parte de los tomadores de decisiones en materia de seguridad de la información (ESET, 2021).

Bajo este hecho, todo ataque informático queda asentado en los logs o registros; sin embargo, surge como interrogante ¿si todo evento malicioso genera una huella en los logs de los activos de información, por qué a menudo pasan desapercibidos?, o en el mejor de los casos, se desconoce el haber sido objeto de un ataque informático. La respuesta a ello está en el volumen de logs que se generan continuamente en las organizaciones.

En consecuencia, la seguridad informática está recibiendo en estos últimos años más atención, debido a que se ha observado y evidenciado, un número creciente de actos de delincuencia informática. La visualización de logs puede ayudar a atenuar este problema. Los logs se utilizan para realizar un seguimiento de todos los usuarios que han accedido a un servidor y podría detectarse comportamientos no regulares (Nadeem y Huang, 2018).

Debido a que cada vez más, la detección de anomalías en los registros de seguridad recibe mayor atención, los eventos de autenticación constituyen un componente importante de estos registros, y con ellos se puede producir predicciones confiables y precisas, que minimicen el esfuerzo de los ciber-expertos para detener ataques falsos (Kaiafas et al., 2018).

Para dilucidar esta problemática, es imprescindible utilizar una herramienta de software que recolecte todos estos datos, los procese bajo reglas definidas que se haya proporcionado y permita su visualización para la interpretación, como por ejemplo un correlacionador de eventos de seguridad o solución que centralice y almacene logs por un periodo extendido de tiempo (Mujawar y Kulkarni, 2015). La visualización de la seguridad se refiere a observar los datos recopilados de una manera diferente, es decir, gráficamente, generar modelos que representen infinidad de datos que resultarían incomprensibles, a primera vista.

La gestión de grandes volúmenes de datos o también conocida como Big Data, hace referencia a sacar el mayor provecho de una enorme cantidad de datos generados por plataformas tecnológicas, sistemas o usuarios (Pérez Marqués, 2015). Comprende el proceso de extracción de información útil a partir de grandes cantidades de datos ofreciendo una visión más amplia de los riesgos y vulnerabilidades, por medio de diversos algoritmos de aprendizaje automático para clasificación y predicción que hacen posible identificar comportamientos anormales mucho antes de que ocurran. Por lo tanto, pueden utilizarse para mejorar la seguridad (Joglekar y Pise, 2016).

Big Data también permite identificar irregularidades y posibles violaciones de la seguridad de la red. Esta tecnología, aparte de ofrecer una gran capacidad a la hora de procesar datos y analizarlos, también permite identificar a las personas que actúen en contra de la integridad de los

mismos. Esto implica que las personas con intenciones de vulnerar la privacidad de los datos se lo piensen dos veces antes de cometer alguna acción que ponga en peligro la información de otras personas (Tounsi y Rais, 2018).

Big Data, entonces puede representar un recurso valioso para el equipo de seguridad de la información, facilitando la automatización de procesos y la detección de anomalías o eventos de seguridad, que normalmente pasarían desapercibidos debido al volumen de datos procesados, a través de la creación de plataformas inteligentes de ciberseguridad concebidas para erradicar las amenazas informáticas. Utilizando herramientas de Big Data se puede iterar rápidamente a través de los datos, construir modelos y proporcionar un rápido análisis visual, facilitando el trabajo para el equipo de ciberseguridad que tiene que analizar millones de logs cada día (Joglekar y Pise, 2016).

De lo antes expuesto, surge esta investigación que tiene como objetivo proponer una arquitectura de análisis de datos a través de herramientas de Big Data mediante la utilización de eventos o registros de seguridad, que permitan mejorar la identificación, integración y correlación de eventos de seguridad. Se plantea el diseño de una arquitectura basada en la integración de Elastic Stack y sus componentes principales (Elasticsearch, Logstash, Kibana), y tecnologías como Filebeat y Wazuh Security Detection (NIPS/HIDS), gestionado la seguridad en activos de información como equipos de comunicaciones, servidores de datos y aplicaciones, motores de bases de datos, y terminales de usuario final.

Marco teórico

Generalidades de Big Data

El descubrimiento de información útil constituye un tema importante entre los investigadores debido a que ésta representa la piedra angular en una sociedad progresista. Se vive hoy en la era de las computadoras e Internet, época en la que los datos se acumulan mediante un crecimiento exponencial a cada segundo. Por lo tanto, es responsabilidad del investigador de ciencias de la computación inventar una mejor manera de obtener información útil de estos datos a gran escala o Big Data (Rohit et al., 2018).

En la actualidad, el uso del término Big Data, tiende a referirse al análisis del comportamiento del usuario, otorgando valor a los datos almacenados y formulando predicciones a través de los patrones de comportamiento observados. Esta disciplina, dedicada al estudio de los datos masivos, se enmarca en el sector de las TIC (Chalmers et al., 2013).

Big Data se ocupa de todas las actividades relacionadas con los sistemas que manipulan ingentes conjuntos de datos, siendo las dificultades más comunes las vinculadas a la gestión de estas cantidades de datos que se centran en la recolección y el almacenamiento, búsqueda, compartición, análisis, y visualización (Hashem et al., 2015). La "gestión de datos masivos", también denominada "inteligencia de datos", "datos a gran escala" o "Big Data", son los términos que hacen referencia al conjuntos de datos tan grandes y complejos, que hacen falta aplicaciones informáticas no tradicionales de procesamiento de datos para tratarlos adecuadamente (Pérez Marqués, 2015).

La tendencia a manipular tales cantidades de datos se debe a la necesidad de incluir dicha información para la creación de informes estadísticos y modelos predictivos utilizados por los directivos de las organizaciones para el análisis de rendimiento de un negocio, marketing y fidelización de clientes, datos de enfermedades infecto-contagiosas, espionaje industrial y seguimiento a los ciudadanos, o en la lucha contra la delincuencia organizada, entre otras actividades (Liu et al., 2014).

El concepto de Big Data se aplica para toda aquella información que no puede ser procesada o analizada, utilizando procesos o herramientas tradicionales; se le consideran seis características claves, que se describen en la tabla 1, la cual plantea la información relativa al Big Data en referencia a la seguridad de la información ((Pérez Marqués, 2015); (Lněnička et al., 2017)).

Tabla 1

Características de Big Data

Característica	Descripción
Volumen	Los datos se producen en mayores cantidades que los datos tradicionales.
Variedad	Distintas tipologías y estructuras de los datos procediendo de fuentes muy diversas. Clasificados en: estructurados, no estructurados y semiestructurados.
Velocidad	El procesamiento de los datos debe hacerse en el menor tiempo posible e incluso en el tiempo real para acceder a los datos, para facilitar el análisis y procesamiento.
Valor	El valor de cualquier dato varía conforme a su contenido se requiere identificar la información valiosa, transformarla y extraer los datos para su posterior análisis.
Veracidad	La información recopilada debe gozar de un alto nivel de fiabilidad, eliminar posibles eventos de inexactitud o incertidumbre.
Viabilidad	Se relaciona con la capacidad de las organizaciones para utilizar de manera eficaz el gran volumen de datos que manejan.

Nota: Basado en Pérez Marqués (2015) y Lněnička et al. (2017).

Seguridad de la información

El término seguridad de la información ha sido acuñado por diferentes autores, confluyendo en la protección de la confidencialidad, integridad y acceso a la información; engloba tecnología, procesos y personas con el propósito de mitigar las amenazas a la información; empleando diferentes medidas técnicas, entre las cuales se destacan software especializado antivirus y antispyware, dispositivos biométricos hasta llegar a los firewalls (Veiga y Eloff, 2007).

El panorama de seguridad moderno está influyendo a los sitios en la red, y en la nube está en constante evolución; trayendo consigo amenazas se visualizan desde una variedad de vías. Es vital desarrollar capacidades de seguridad operativa en todo contexto mundial (Crooks y Vâlsan, 2019). La seguridad de la información comprende aquellos procesos, buenas prácticas y metodologías que aspiran la protección de la información y los accesos a los sistemas de información, su uso, divulgación, interrupción, modificación o destrucción no autorizada (ISO / IEC, 2014). También se relaciona con los procesos de registro y baja de usuarios y aspectos

personales de los usuarios. Todos estos aspectos, deben tenerse en cuenta cuando se despliega la seguridad de la información. ((Kritzinger y Smith, 2008); (Veiga y Eloff, 2007)).

Aspectos legales que dan sustento a la propuesta

La seguridad de la información es un concepto que permite mejorar los procesos inherentes a los sistemas de información y asegurar su nivel de protección, minimizando los riesgos propios de su funcionamiento. Desde la Constitución del Ecuador y diferentes leyes orgánicas, se da sustento legal al desarrollo de técnicas, utilizadas para salvaguardar datos contenidos en los sistemas de información.

En el artículo 16, numeral 2 de la Constitución de la República del Ecuador (2008), se señala que todas las personas, en forma individual o colectiva, tienen derecho al acceso universal a las tecnologías de información y comunicación. Y, así mismo, en el artículo 66, numeral 19 se indica que, se reconoce y garantizará a las personas el derecho a la protección de datos de carácter personal, que incluye el acceso y la decisión sobre información y datos de este carácter, así como su correspondiente protección.

Por otro lado, el artículo 76 de la Ley Orgánica de Telecomunicaciones (2015), plantea sobre las medidas técnicas de seguridad e invulnerabilidad, que las y los prestadores de servicios ya sea que usen red propia o la de un tercero, deberán adoptar las medidas técnicas y de gestión adecuadas para preservar la seguridad de sus servicios y la invulnerabilidad de la red y garantizar el secreto de las comunicaciones y de la información transmitida por sus redes. Dichas medidas garantizarán un nivel de seguridad adecuado al riesgo existente.

Otra muestra del marco legal que soporta la seguridad de información se señala en el artículo 10 de la Ley Orgánica de Protección de Datos Personales(2021), sobre el principio de seguridad de datos personales, en el que se manifiesta que los responsables y encargados de tratamiento de los datos personales deberán implementar todas las medidas de seguridad adecuadas y necesarias, entendiéndose por tales las aceptadas por el estado de la técnica, sean estas organizativas, técnicas o de cualquier otra índole, para proteger los datos personales frente a cualquier riesgo, amenaza, vulnerabilidad, atendiendo a la naturaleza de los datos de carácter personal, al ámbito y el contexto.

Aplicaciones de herramientas de Big Data a nivel de seguridad

La Big Data se relaciona con las áreas de conocimiento de seguridad informática, seguridad de red de datos, y seguridad de la información, dado que los datos generados por las distintas fuentes de información (internas o externas) deben estar protegidos ante eventos de carácter malicioso o mal intencionado, como son los ataques o amenazas informáticas (Joglekar y Pise, 2016).

La protección y la conservación de la integridad de las fuentes de datos masivas, requiere de expertos especializados en seguridad de la información y análisis de datos, dado que las

amenazas informáticas continuamente se diversifican y hacen cada vez más compleja la tarea de mantener una seguridad de extremo a extremo (Roji y Sharma, 2019). Esto se debe a que estudios han demostrado que los ataques avanzados de amenazas persistentes son difíciles de detectar utilizando enfoques tradicionales de detección de intrusos basados en anomalías (Chacon et al., 2020). Big Data y sus herramientas de ecosistema son aplicadas en el análisis, gestión y predicción de eventos de seguridad, detectando y pronosticando posibles amenazas en tiempo real o de manera histórica (Cortés et al., 2017).

Sistema de gestión de información y eventos de seguridad

La gestión y el tratamiento de las amenazas informáticas, es uno de los aspectos más importantes a considerar por parte de las organizaciones modernas, dedicando los recursos necesarios para poder responder a cualquier incidente de seguridad que surja. El sistema de gestión de información y eventos de seguridad (SIEM por sus siglas en inglés Security Information and Event Management Systems) actúa como un repositorio o contenedor de datos, centralizado con bases de datos de gran tamaño para almacenar y gestionar los datos de operación, registrando todos los eventos de los activos de información (Kumar et al., 2018); que se relacionan con la seguridad, usándolos para monitorear, identificar, documentar e incluso responder a dichos incidentes de seguridad (Labrinidis y Jagadish, 2012).

El SIEM comprende la supervisión en tiempo real junto con el análisis de los datos de registro para que la correlación de eventos puede proporcionar la detección, identificación y notificación de anomalías (Kumar et al., 2018). Algunos de los incidentes de seguridad son claros y su identificación podría ser sencilla, pero una gran parte de los incidentes de seguridad, aunque puedan ser obvios, se esconden detrás de la gran cantidad de eventos por segundo que se producen en los activos de información de la organización, y que sin un SIEM serían completa o parcialmente inadvertidos.

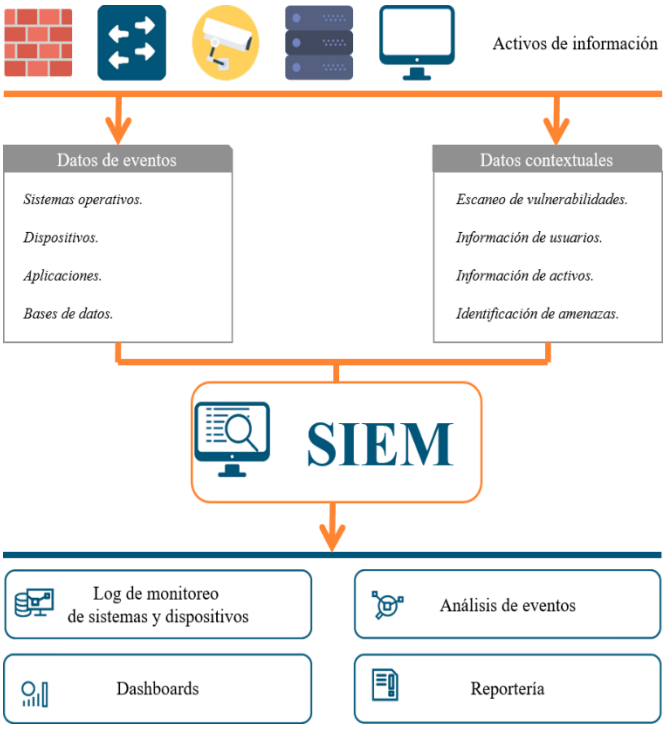
De igual manera, podrían existir dificultades o problemas identificados al momento de implementar la mayoría de SIEM de alto rendimiento, pues son muy costosos y están integrados en una caja negra; pocas organizaciones los adquirirían. El diseño e implementación de un SIEM de código abierto con las mismas funcionalidades que los comerciales es un desafío científico. Los SIEM clásicos actuales no gestionan Big Data, trae consigo que el análisis de datos de seguridad de fuentes heterogéneas puede resultar difícil para la detección de intrusiones cuando las fuentes homogéneas ya tienen problemas con grandes cantidades de estos datos (Arass y Souissi, 2019).

Arquitectura referencial del SIEM

La arquitectura referencial propuesta, de un SIEM, se conforma por una estructura tecnológica compleja, descrita en la figura 1; cuyos componentes se interconectan de forma física y lógica a partir de sus funciones clave.

Figura 1

Arquitectura referencia de SIEM



Fases de Big Data

Los diferentes tipos de análisis de seguridad para la identificación de amenazas informáticas siguen un conjunto de fases comunes que permiten la toma de decisiones. Las fases de procesamiento de Big Data se consideran como una serie de pasos que deben seguir las organizaciones durante la gestión de la información (Labrinidis y Jagadish, 2012).

- Adquisición, registro
- Extracción, limpieza, anotación
- Integración, agregación y representación
- Análisis y modelado
- Interpretación

Componentes de Elastic Stack

Elastic stack es una plataforma de código abierto para la ingesta confiable de datos de diferentes fuentes, en una variedad de formatos distintos; que permite buscar, analizar y visualizar datos en tiempo real (Crooks y Vâlsan, 2019). Es utilizada para crear soluciones de Big Data, está compuesta por: Elasticsearch (ES), Logstash y Kibana (Talas et al., 2017). Se conoce como un motor de búsqueda y analítica de tipo código abierto, basado en Apache Lucene, considerado como un buscador de texto completo, distribuido y con capacidad de multi-tenencia con una interfaz web RESTful (Basado en arquitectura REST, que es una interfaz para conectar varios sistemas basados

en HTTP, y sirve para obtener y generar datos y operaciones) y con documentos JSON, diseñado para permitir una escalabilidad horizontal, fiabilidad y de fácil gestión.

Elastic Stack como solución tecnológica puede utilizarse para la visualización y el análisis no solo de los registros de fallas del sistema, sino también de los registros de tareas habituales en los sistemas comerciales (Maeda et al., 2018).

El cuadrante mágico de Garner incluye en el mes de febrero 2021 a Elastic, descrito en la figura 2, dentro de los principales quince proveedores de motor de información que combina las capacidades de búsqueda con la inteligencia artificial, para ofrecer información procesable derivada del espectro completo de contenido y datos, obtenidos dentro y fuera de una empresa, ayudando de esta manera a los líderes de aplicaciones a tomar la mejor decisión (Gartner, 2021).

Figura 2

Cuadrante Mágico de Garner para motores Insight

Nota: Extraído de Gartner, Inc (2021)

Se combinan las tres soluciones (Elastic search, Logstash, Kibana) para construir una solución de Big Data (Talas et al., 2017), según las especificaciones de la tabla 2:

Tabla 2

Especificaciones de las soluciones de Elastic Stack

Solución tecnológica	Descripción
Elastic search (ES)	Permite la búsqueda de texto completo en datos no estructurados no depende del tipo de fuente de datos: estructurada, semiestructurada o estructurados o no estructurados.
Logstash	Se utiliza para recoger y analizar los datos en un servicio central y enviar el resultado a ES para su indexación. Se puede personalizar y adaptarse para procesar cualquier fuente de datos como registros, netflow, bases de datos bases de datos, archivos, etc.
Kibana	Constituye la interfaz gráfica de la solución que muestra y busca los datos de ES. Contiene una sintaxis de búsqueda para consultar el ES. Facilita la creación de cuadros de mando interactivos, pues proporciona diversos tipos de gráficos predefinidos como pasteles, histogramas o tendencias.

En la actualidad, las amenazas cibernéticas se vuelven cada vez más sofisticadas, se necesita del análisis de seguridad y del monitoreo en tiempo real, para una rápida detección y reparación de amenazas. El sistema de detección de intrusos Wazuh forma parte del SIEM, por lo tanto, se integra con Elastic Stack. Se conoce como una plataforma de código abierto que se utiliza para recopilar, agregar, indexar y analizar datos de seguridad, lo que ayuda a las organizaciones a detectar intrusiones, amenazas y anomalías de comportamiento (wazuh, 2021b).

Entre tanto, el componente Filebeat se convierte en un reenviador liviano que se usa para transmitir logs a través de una red, generalmente a ES. Se utiliza en el servidor de Wazuh para enviar eventos y alertas a ES. Se encarga de leer la salida del motor de análisis de Wazuh y envía eventos en tiempo real a través de un canal encriptado (wazuh, 2021a).

Metodología

Para el desarrollo del presente trabajo se siguió un método de investigación de carácter exploratorio, descriptivo, generando valor junto con la experiencia y conocimiento de expertos especializados en seguridad de la información y análisis de datos, residiendo su aplicabilidad en el entorno de desarrollo de arquitecturas tecnológicas, utilizando herramientas del ecosistema de Big Data para mejorar la identificación de amenazas informáticas.

Dentro de los procesos llevados a cabo en la investigación descritos en la figura 3, se planteó un escenario controlado que simule un Centro de Operaciones y Monitoreo de Seguridad de la Información de una empresa que brinda servicios tecnológicos en el Ecuador, con carácter práctico-investigativo y. con la finalidad de analizar y evaluar posibles brechas de seguridad de la información, sobre la base del diseño de una arquitectura tecnológica, de procesos, de software, de negocios, sustentando de esta manera las fases del proceso de Big Data propuestas por (Labrinidis y Jagadish, 2012) para la identificación de amenazas informáticas.

Figura 3

Proceso de diseño de arquitectura tecnológica

Adquisición y registro

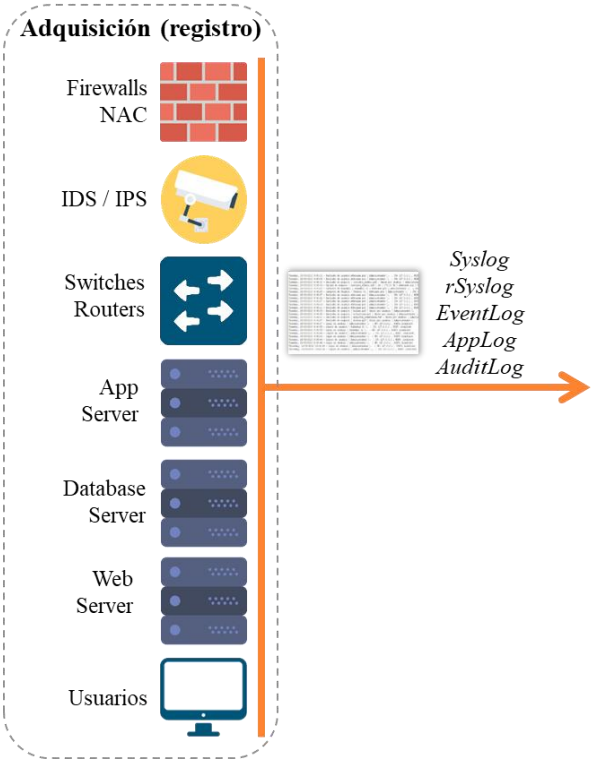
Esta etapa comprende el primer paso del procesamiento, incluye el abastecimiento de diversas fuentes de datos: servidores de datos, aplicaciones, terminales de usuarios, equipos de seguridad, redes de sensores, entre otros los cuales pueden producir cantidades asombrosas de datos en bruto.

Se recopilaron los datos con información potencialmente útil para el análisis de eventos de seguridad de la información, desde los activos de información como servidores de datos o aplicaciones, terminales de usuarios, equipos de seguridad, equipos de comunicaciones, entre otros, en formato de eventos (logs).

En esta fase, se utiliza los agentes Filebeat, como plataforma para los gestores de datos y posterior envío desde los activos de información hacia procesos posteriores (Logstash y ES), ver la figura 4. Este aplicativo ligero permite reenviar y centralizar los logs; se instala como un agente en sus servidores. Filebeat monitorea los archivos logs o las ubicaciones que se requieran, recopila eventos de log y los reenvía a ES o Logstash para su indexación (Elastic, 2021).

Figura 4

Fase adquisición y registro

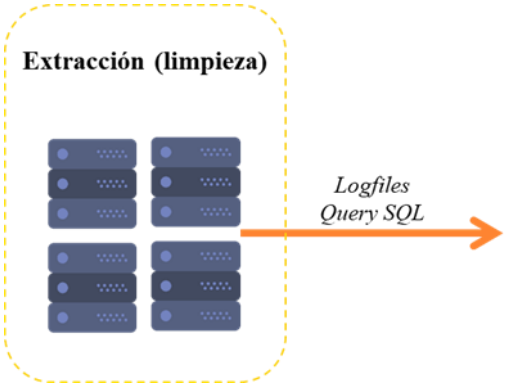


Extracción, limpieza, anotación

En esta fase (figura 5) se extrae la información necesaria de las fuentes subyacentes y la expresa en una forma estructurada, adecuada para el análisis. Por intermedio de Wazuh se procede a realizar la detección de intrusos basado en un host de código abierto (HIDS), proporcionando el análisis de registros, supervisión de integridad de archivos de sistema operativo, detección de rootkits y vulnerabilidades, evaluación de configuraciones y capacidad de respuesta ante incidentes de seguridad de la información.

Figura 5

Fase extracción, limpieza y anotación



Esta solución integral se integra a las herramientas adicionales como OpenSCAP y ES, tal como se describe en las figuras 6 y 7. Mientras que, Logstash, actúa como canal de procesamiento de datos de código abierto en el lado del servidor de Big Data, que ingiere datos de una multitud de fuentes simultáneamente, los transforma y envía para su almacenamiento.

Figura 6

Interfaz gráfica, Gestión de Agentes desplegados

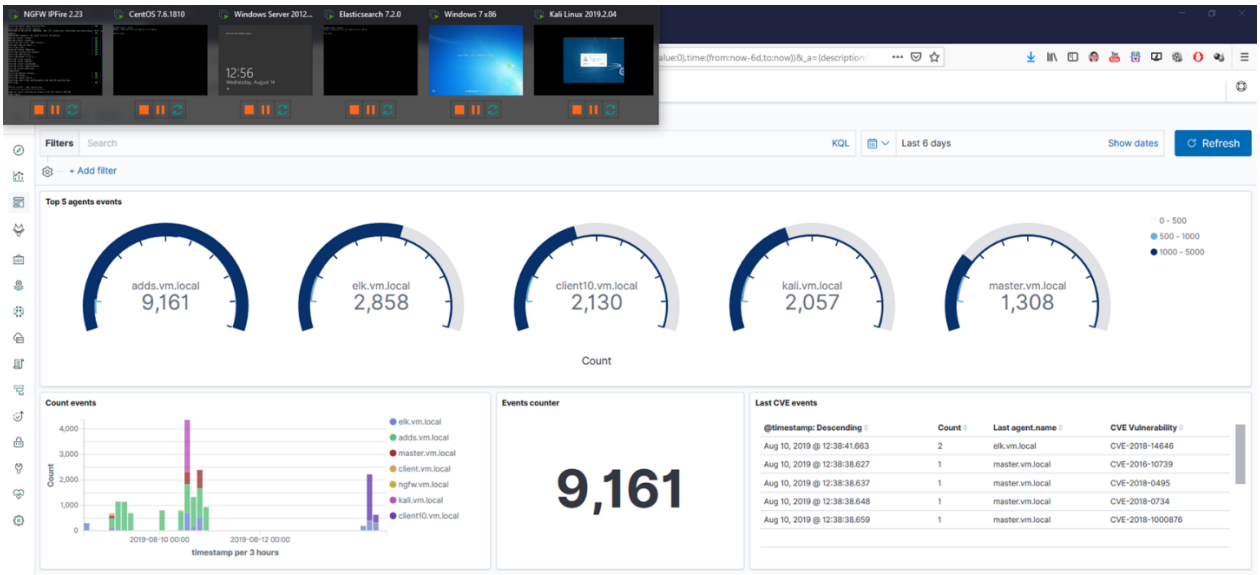
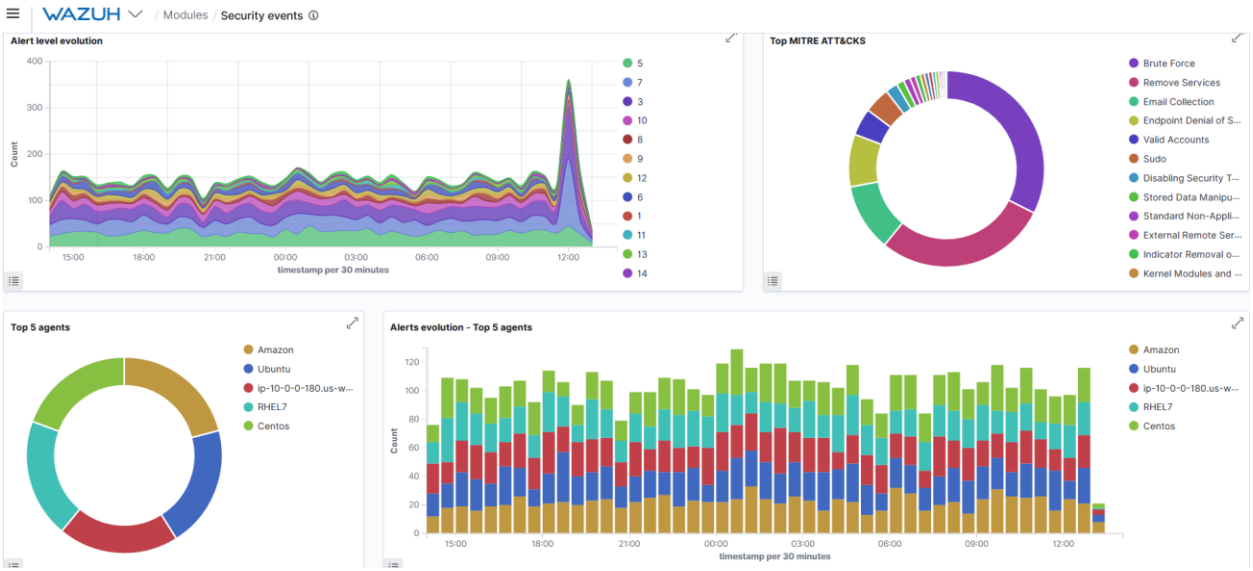


Figura 7

Interfaz gráfica, Perspectiva general para identificación de amenazas

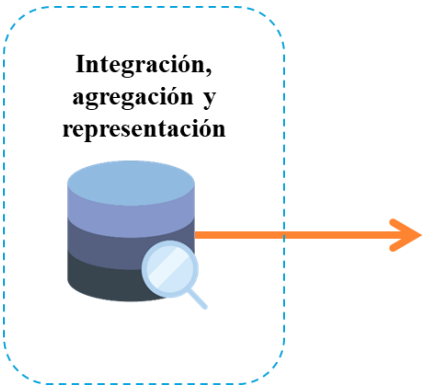


Integración, agregación y representación

Esta fase (figura 8) tiene como principal función el procesamiento y transformación de los datos, que se puede almacenar y mostrar de manera más entendible. La agregación de datos se realizó a través de un software especializado, por el ES, como motor de búsqueda y análisis, capaz de adaptarse a un número creciente de casos de uso.

Figura 8

Fase integración, agregación y representación

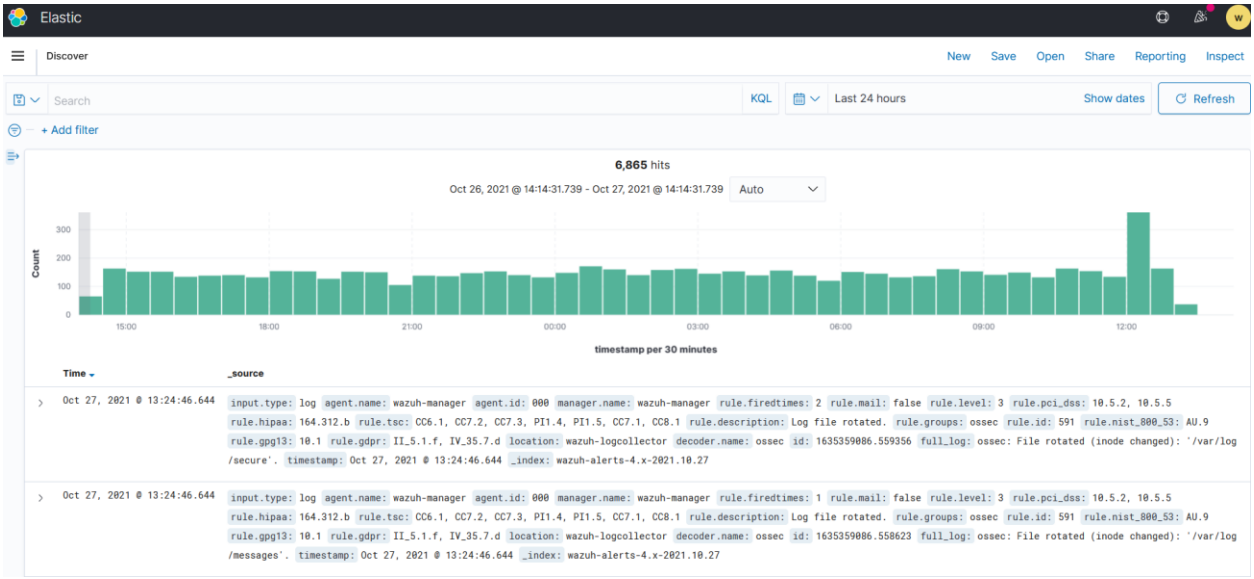


El software en mención se encarga de almacenar los datos de forma centralizada y distribuida, permitiendo ejecutar búsquedas avanzadas. Por medio de la interfaz gráfica de Elastic Stack, descrita en la figura 9, se realiza el seguimiento de registros en tiempo real, a través de una

pantalla unificada y personalizable. Los datos de registros se correlacionan con las métricas de la interfaz de usuario de infraestructura, lo que facilita el diagnóstico de los posibles incidentes.

Figura 9

Interfaz gráfica de Elastic Stack



Análisis y modelado

Dentro del análisis y modelado de Big Data (figura 10), se consideran como herramientas al conjunto de aplicaciones, componentes y normas que procesan grandes cantidades de datos, siendo sus resultados, elementos claves para la toma de decisiones.

Figura 10

Fase análisis y modelado

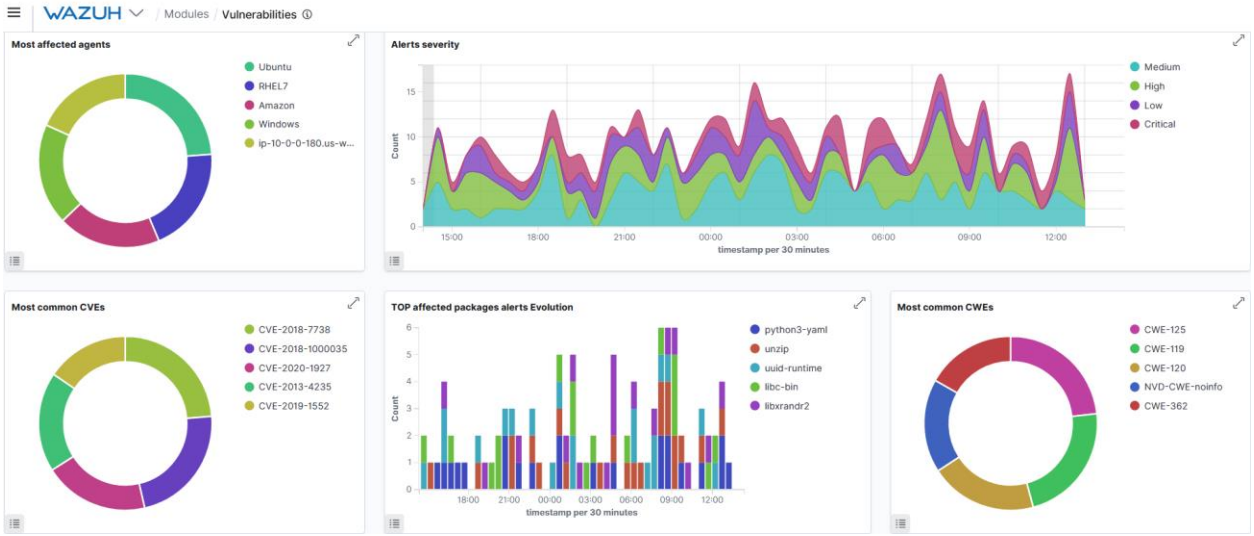
La tecnología utilizada en el análisis fue Kibana, interfaz de usuario en el Elastic Stack. Facilitó el análisis de los eventos basados en el tiempo a través de la visualización. Con la creación de paneles interactivos, permitió una mayor comprensión y visibilidad.

El índice creado en Elastic Search se utiliza en Kibana para analizar y crear visualizaciones. También contempla la detección de anomalías, alertas y el monitoreo basados en el aprendizaje automático (P. y C., 2019).

En la figura 11, se describe el uso de Kibana con la visualización de los datos de las búsquedas ejecutadas en ES, y navegar por el Elastic Stack a nivel de Dashboards generados, mostrando toda la información que se tiene almacenada e indexada como datos no estructurados (Lněnička et al., 2017).

Figura 11

Interfaz gráfica, Dashboard de seguridad



Interpretación

Esta fase se apoya de la herramienta Kibana, con sus dashboards generados permitiendo desde el seguimiento de la carga de consultas hasta la comprensión de solicitudes de trabajo que fluyen a través de las aplicaciones.

Resultados

Una vez realizado el inventario de datos y en conformidad con el propósito de la investigación, se procedió a la evaluación de brechas de seguridad de la información, aplicando una arquitectura tecnológica-empresarial que permitió la implementación de herramientas del ecosistema de Big Data, para la identificación de amenazas informáticas descritas e ilustradas en la tabla 3 y figura 12.

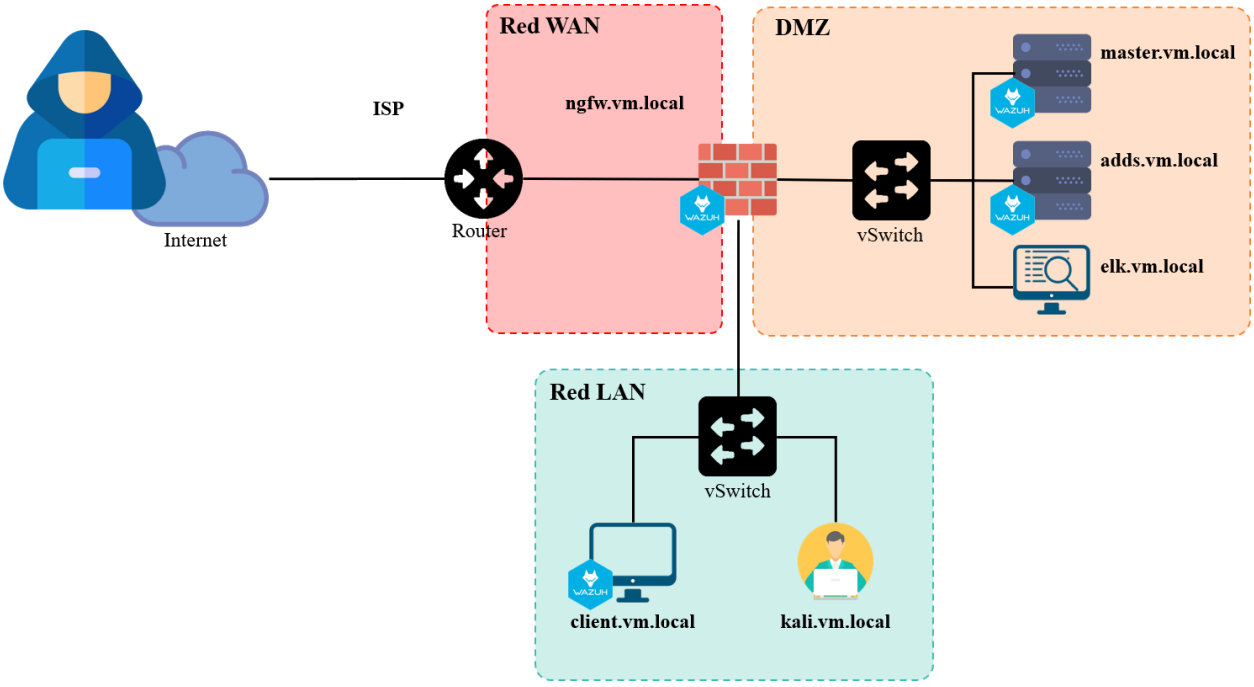
Tabla 3

Consolidado de infraestructura virtual implementada

Host	Máquina virtual	Nombre del servidor	Descripción
Host 1	VM1	ngfw.vm.local	Next-gen firewall, DNS, DHCP, filtro de URL, filtro de contenido, IDS/IPS.
	VM2	master.vm.local	DNS, servidor Web, servidor DB, servidor FTP, servidor NTP.
	VM3	adds.vm.local	Servidor de autenticación (ADDS), servidor DB, servidor DLP.
	VM4	elk.vm.local	Elasticsearch, Logstash, Kibana, Wazuh Manager.
	VM5	client.vm.local	Cliente de dominio.
	VM6	kali.vm.local	Monitor de explotación de vulnerabilidades.

Figura 12

Arquitectura de infraestructura tecnológica



Para efecto de la ejecución de las pruebas de validación de la infraestructura propuesta se aplicaron cinco casos de uso de ataques informáticos simulados, bajo un ambiente controlado, sobre la infraestructura tecnológica desplegada como parte de la presente investigación, los cuales han sido desarrollados para la explotación de las vulnerabilidades informáticas más comunes.

Caso de uso para validación de resultados

En la presentación de los casos se evalúa el proceso de análisis y monitoreo de eventos, a partir del consumo de la información, obtenida mediante la gestión de los activos de información

ISSN 2477-9024. Innova Research Journal (Septiembre-Diciembre, 2021). Vol. 6, No. 3.2, pp. 141-167

de la organización o fuentes generadoras de información con relación a eventos de seguridad los Logs.

Caso de uso No. 01.

El usuario promedio emplea al menos cinco contraseñas diferentes que va intercambiando entre las distintas aplicaciones corporativas, dispositivos electrónicos (terminal o laptop de usuario final, tablet o móvil) y cuentas transaccionales, como se describe en la tabla 4.

Tabla 4

Caso de uso No. 01.

CUSEG No. 01	**Violación de políticas de acceso y uso de credenciales de acceso**
Descripción	Controlar el acceso autorizado a los sistemas y aplicaciones de la organización desde una o varias ubicaciones geográficas.
Objetivo	Detectar incumplimiento de políticas de control de acceso de usuarios.
Condiciones de alerta	a. Alertar sobre múltiples autenticaciones simultáneas de un mismo usuario. b. Inicio de sesión desde ubicaciones geográficas no autorizadas. c. Alertar sobre intentos fallidos de autenticación o conexión.
Fuentes de datos	Logs del sistema de autenticación de usuarios (ADDS). Firewall. IPS / HIDS. Proxy Server.

En la figura 6, se visualizan los agentes desplegados en la infraestructura tecnológica implementada durante la investigación, identificando que el servidor *ngfw.vm.local* presenta alertas de seguridad por utilización recurrente de credenciales de acceso a los servicios corporativos. Por su parte, la figura 13 presenta un resumen descriptivo de las alertas de seguridad presentadas conforme a la clasificación propuesta por MITRE Corporation en su clasificación ataques informáticos más comunes.

Figura 13

Visualización de eventos de seguridad, según MITRE ATT&CKS

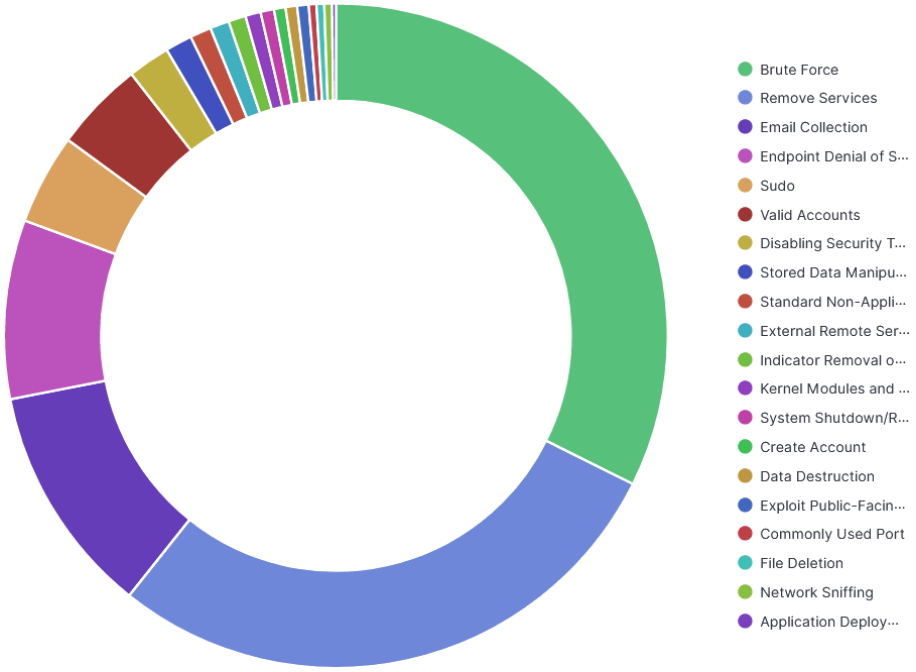


Figura 14

Lista resumen de alertas de seguridad

Security Alerts

Time ↓	Agent	Agent name	Technique(s)	Tactic(s)	Description	Level	Rule ID
Oct 27, 2021 @ 13:24:46.644	000	wazuh-manager			Log file rotated.	3	591
Oct 27, 2021 @ 13:24:46.644	000	wazuh-manager			Log file rotated.	3	591
Oct 27, 2021 @ 13:24:46.644	000	wazuh-manager			Log file rotated.	3	591
Oct 27, 2021 @ 13:07:17.894	006	Windows			CVE-2019-18684 affects sudo	7	23504
Oct 27, 2021 @ 13:07:06.474	005	Centos			Windows: Service startup type was changed.	3	1873
Oct 27, 2021 @ 13:06:58.888	001	RHEL7	T1021	Lateral Movement	sshd: Possible attack on the ssh server (or version gathering).	8	5701
Oct 27, 2021 @ 13:06:19.718	006	Windows			Sample alert 5	2	4026

Caso de uso No. 02.

En el campo de la seguridad de la información, la infraestructura de centro de comando y control (Command and control, usualmente abreviado C&C o C2), consta de servidores de

procesamiento y otros elementos que son usados para controlar los malware y botnet, tal como se presenta en la tabla 5.

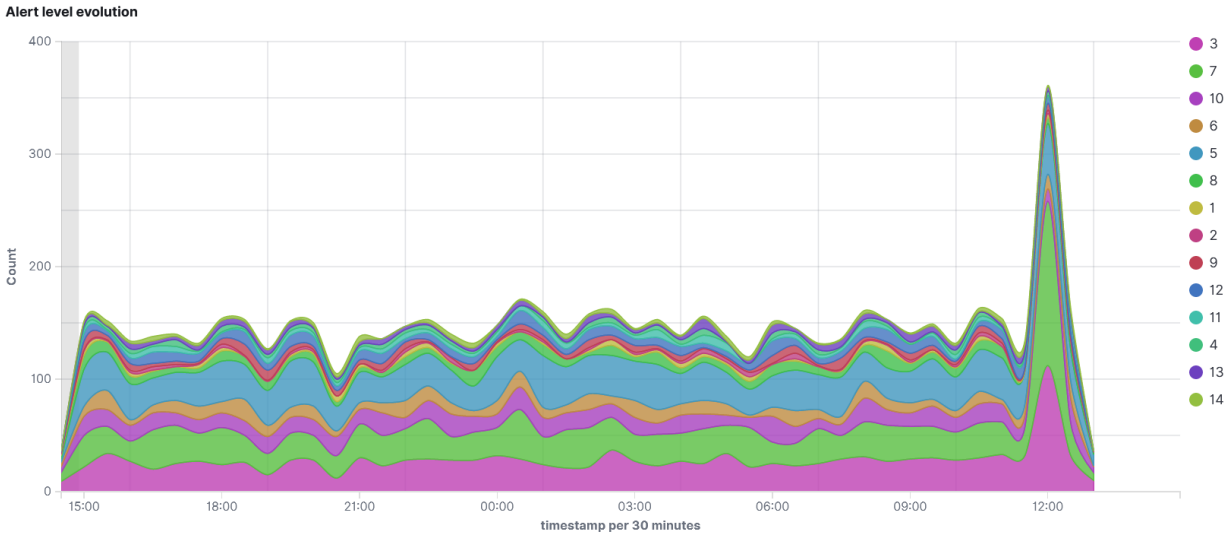
Tabla 5
Caso de uso No. 02.

CUSEG No. 02	**Detección eventos salida Internet (Comando de centro y control)**
Descripción	Cuando un equipo interno (terminal de usuario) de la organización genere un alto número de peticiones hacia Internet, este indicará en el equipo de seguridad perimetral y de prevención de intrusos (IPS) que es posible que una máquina esté contaminada con malware o esté enviando ataques hacia fuentes externas.
Objetivo	Detectar incumplimiento de políticas de control de acceso de usuarios.
Condiciones de alerta	Alertar cuando ocurran los siguientes eventos de forma simultánea: a. Múltiples conexiones hacia redes externas. b. Detección de firma que involucre la misma dirección IP destino.
Fuentes de datos	Firewall. IPS / HIDS.

En una organización, el equipo de seguridad necesita encontrar una manera eficiente de detectar y controlar aspectos relevantes a las conexiones desde y hacia Internet para mantener un entorno seguro y accesible. Se muestra en la figura 15, una de las formas de obtener el control por parte de los atacantes es a través del protocolo DNS, siendo las brechas de seguridad de Windows el principal medio de utilización para este tipo de ataque o amenaza informática. Aplicando las contramedidas y contención través de estos servicios, se puede interrumpir la transmisión de paquetes de un malware hacia el punto de conexión o servidor de comando y control (C&C).

Figura 15

Visualización de evolución del nivel de alertamiento de seguridad



Caso de uso No. 03.

Además de los ataques de código malicioso (virus, troyanos, spyware, malware, entre otros), es posible que los activos de información de una organización sean objetivo de diversos ataques de red de datos, uno de los más comunes es el de reconocimiento, detección y esquematización no autorizadas de sistemas, servicios o vulnerabilidades, ver la tabla 6.

Tabla 6

Caso de uso No. 03.

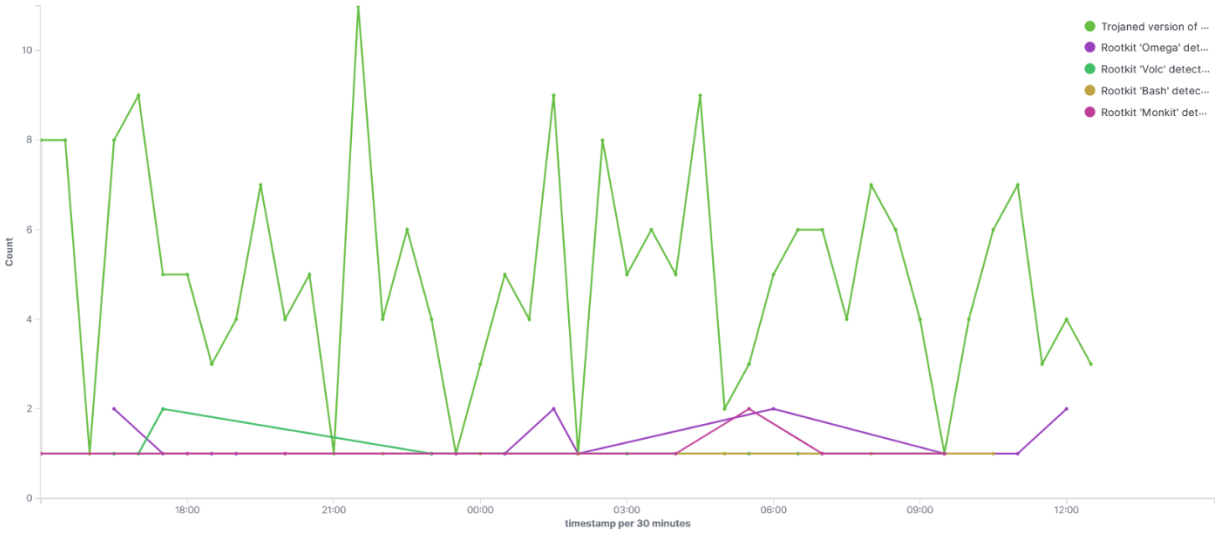
CUSEG No. 03	**Detección reconocimiento objetivos**
Descripción	Reconocimiento de servicios críticos.
Objetivo	Detectar un posible ataque interno de reconocimiento de puertos a los servidores principales de la red de datos organizacional.
Condiciones de alerta	Alertar cuando ocurran los siguientes eventos de forma simultánea: a. Detecta conexiones fallidas y satisfactorias hacia los diferentes servidores corporativos. b. Se detectan firmas relacionadas con reconocimiento hacia la granja de servidores con destino de direcciones críticas para la organización.
Fuentes de datos	Firewall. IPS / HIDS. Topología de la infraestructura.

El escaneo de activos intenta encontrar potenciales incidentes de seguridad usando ataques conocidos contra objetivos seleccionados. Como se observa en la figura 16, se verifica el escaneo de activos de información sobre los objetivos identificados, que para la investigación corresponde a la infraestructura tecnológica desplegada.

Cabe destacar que, el escaneo de activos solo puede encontrar ciertos tipos de incidentes de seguridad y vulnerabilidades lógicas, tales como un control de acceso defectuoso. Se verifica, además, la ejecución de una prueba de penetración automática para encontrar incidentes de seguridad relacionados con los cumplimientos normativos de los activos de información de la organización.

Figura 16

Visualización de eventos por tipo de control



Caso de uso No. 04.

Un ataque dirigido es un proceso a largo plazo que pueden comprometer la seguridad y dar acceso no autorizado a los criminales cibernéticos; algunos de estos ataques utilizan amenazas persistentes (APT) o pueden usar técnicas como malware avanzado o exploits de día cero (Zero-day), como se puntualiza en la tabla 7.

Tabla 7

Caso de uso No. 04.

CUSEG No. 04	**Detección de ataques específicos o ataques dirigidos**
Descripción	Alto número de conexiones hacia un servicio especifico
Objetivo	Identificar ataques internos realizados hacia los servidores principales de la organización.
Condiciones de alerta	Alertar cuando ocurran los siguientes eventos de forma simultánea: a. Detecta conexiones satisfactorias hacia los servidores críticos de la organización. b. Se detecta una firma de un servicio Web, consola remota SSH, E-mail, base de datos, etc., con destinos de direcciones críticas para la organización.
Fuentes de datos	Firewalls. IPS / HIDS. Topología de la infraestructura.

Para la ejecución de ataques dirigidos se propone la utilización de la herramienta UFOnet junto con GNU / Linux Kali, tal como se observa en las figuras 17 y 18. UFONet se conoce como una herramienta de código abierto con licencia GPLv3, escrita por Lord Epsylon (psy) en python, html5 y javascript, para realizar ataques de denegación distribuidos (DDoS) aprovechando fallos de redirecciones abiertas (Open redirect) en aplicaciones web a modo de botnet.

Figura 17

Ejecución de un ataque dirigido con UFONet.

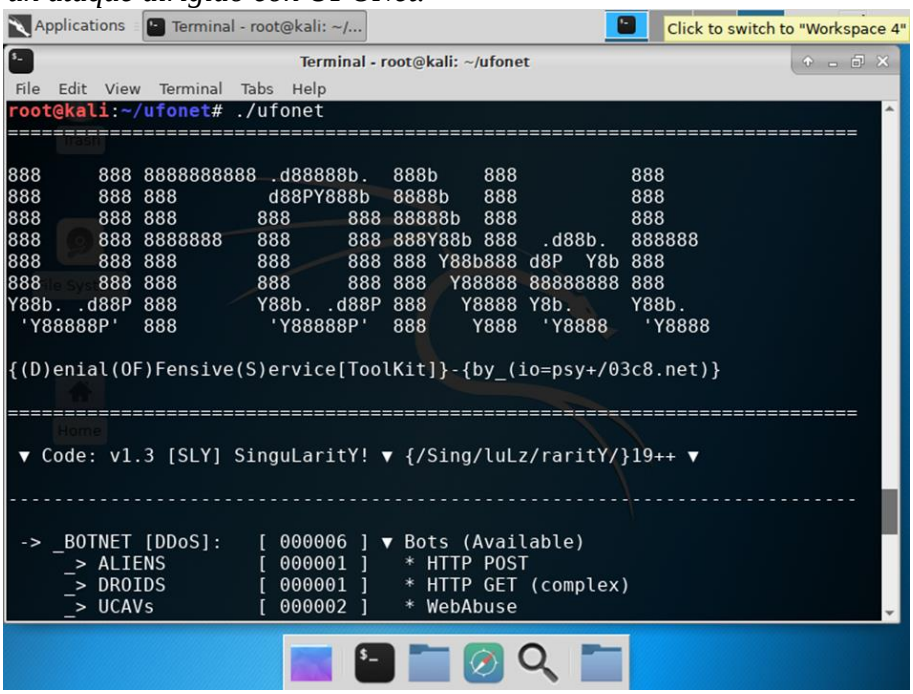
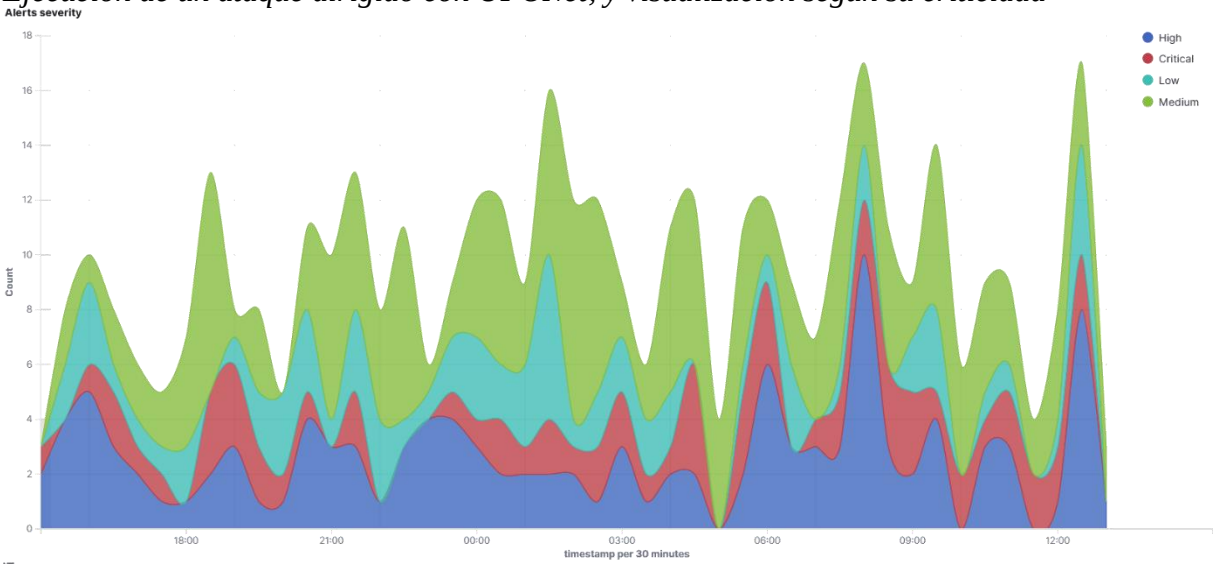


Figura 18

Ejecución de un ataque dirigido con UFONet, y visualización según su criticidad



Caso de uso No. 05.

Las amenazas que se originan afuera de la red de datos corporativa, al no tener información certera de la topología de la red interna de la organización, hacen que el criminal cibernético deba

realizar varias acciones para conocer la estructura y encontrar la manera de atacar los activos de información, ver tabla 8.

Tabla 8

Caso de uso No. 05.

CUSEG No. 05	**Detección de amenazas externas sobre infraestructura crítica**
Descripción	Detectar actividades anómalas provenientes de fuentes externas que puedan afectar la disponibilidad de las aplicaciones críticas de la organización.
Objetivo	Identificar ataques que puedan afectar la disponibilidad de las aplicaciones críticas.
Condiciones de alerta	a. Flujos de tráfico desde una misma dirección IP, desde o hacia uno o varios puertos de servicios. b. Alto número de peticiones originadas desde una misma dirección IP origen, hacia un puerto de servicio específico, como por ejemplo 80 (HTTP) o 443 (HTTPS).
Fuentes de datos	Firewalls IPS / HIDS. Topología de la infraestructura.

Un inventario de activos se define como una lista de todos aquellos recursos (físicos, software, documentos, servicios, personas, instalaciones, entre otros) que contengan valor para la organización y necesiten por tanto ser protegidos de potenciales amenazas informáticas. Como se observa en las figuras 19 y 20, se aplican técnicas de enumeración de activos de información, los cuales facilitan la identificación de los activos de información expuestos sobre la infraestructura tecnológica empleada en la investigación.

Figura 19

Ataque de enumeración de activos de información

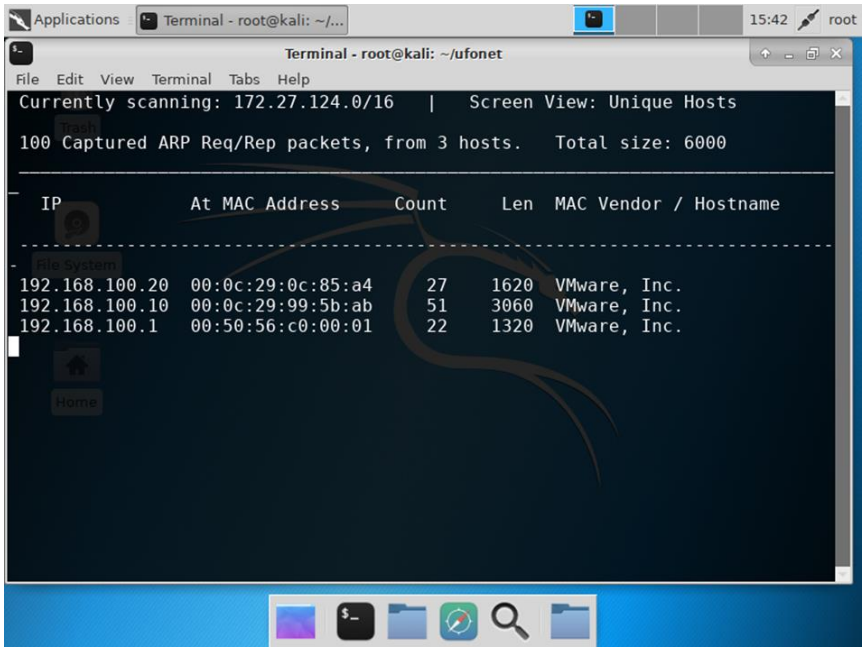


Figura 20

Identificación de activos de información y cuentas de usuarios comprometidas

Top 5 users

Agent ID	Agent name	Top user	Count
002	master.vm.local	root	176
001	ngfw.vm.local	root	151
004	democlient.vm.local	Administrators	1
003	adds.vm.local	Administrators	1

Conclusiones

La tecnología continuará evolucionando, las organizaciones seguirán siendo cada vez más dependientes de las TIC, por consecuencia las amenazas relacionadas a dichos avances se mantendrán e inclusive podrán aumentar debido a configuraciones técnicas deficientes, la inadecuada gestión o la falta de capacidades y competencias técnicas de los proveedores de las tecnologías. El análisis de grandes cantidades de datos en el área de la seguridad de la información es considerado como un área de trabajo reciente, y requiere de una amplia investigación, así como establecer una arquitectura tecnológica que ofrezca una alta capacidad de almacenamiento, flexibilidad técnica y con una inversión de capital menos costosa, para hacer frente a las amenazas informáticas, con el apoyo del análisis de datos con herramientas de Big Data.

En el proceso de aplicación de Big Data es necesario iniciar con la comprensión y conceptualización de cómo se producen los datos en los activos de información de una organización y cuál es el ciclo de vida de estos, a fin de presentar una propuesta para la recopilación, almacenamiento y procesamiento eficaz de grandes volúmenes de datos de seguridad en tiempo real, así como flujo de datos históricos.

Adicionalmente, es importante tener en cuenta el rendimiento y recursos tecnológicos necesarios, para que la plataforma de identificación de amenazas informáticas no sólo opere bajo un esquema de laboratorio de pruebas y ambiente controlado, sino que también se permita su operación en un entorno de producción real. Para la validación e implementación de la arquitectura tecnológica propuesta, que permita la identificación de amenazas informáticas, es primordial definir una guía de recursos de hardware y software, el dimensionamiento de un sistema de almacenamiento distribuido de Elastic Stack, la capacitación del personal de seguridad de la información, además del uso de herramientas de Big Data.

La presente investigación, implementó y validó una arquitectura para la identificación de amenazas informáticas, utilizando herramientas de Big Data, especificando cada uno de los componentes y módulos que intervinieron en cada una de las fases. En virtud del estudio efectuado,

se puede concluir que es imperativo que las organizaciones sigan trabajando en procura de madurar sus modelos de seguridad de la información, fortaleciendo sus equipos de especialistas e implementando nuevos esquemas para la identificación de amenazas informáticas aplicando modelos predictivos.

Referencias bibliográficas

Arass, M. E., & Souissi, N. (2019). *Smart SIEM: From Big Data Logs and Events To Smart Data Alerts*. *8*(8), 6.

Chacon, J., McKeown, S., & Macfarlane, R. (2020). Towards Identifying Human Actions, Intent, and Severity of APT Attacks Applying Deception Techniques—An Experiment. *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 1-8. https://doi.org/10.1109/CyberSecurity49315.2020.9138859

Chalmers, S., Bothorel, C., y Picot-Clemente, R. (2013). *Big Data—State of the Art*. 24.

Constitución de la República del Ecuador. (2008). *Constitución de la República del Ecuador*.

Cortés, C. B. Y., Landeta, J. M. I., y Chacón, J. G. B. (2017). *El Entorno de la Industria 4.0: Implicaciones y Perspectivas Futuras*. 19.

Crooks, D., & Vâlsan, L. (2019). Building a minimum viable Security Operations Centre for the modern grid environment. *Proceedings of International Symposium on Grids & Clouds 2019 — PoS(ISGC2019)*, 010. https://doi.org/10.22323/1.351.0010

Elastic. (2021). *Filebeat overview | Filebeat Reference [7.13] | Elastic*. Filebeat overview. https://www.elastic.co/guide/en/beats/filebeat/current/filebeat-overview.html

ESET. (2021). *S3CURITY R3PORT Latinoamérica 2021*. https://www.welivesecurity.com/wp-content/uploads/2021/06/ESET-security-report-LATAM2021.pdf

Gartner. (2021). *Gartner Reprint*. https://www.gartner.com/doc/reprints?id=1-25H9R6TE&ct=210318&st=sb

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, 98-115. https://doi.org/10.1016/j.is.2014.07.006

ISO / IEC. (2014). *SO / IEC 27000. Tecnología de la información—Técnicas de seguridad—Sistemas de gestión de seguridad de la información—Descripción general y vocabulario*. https://www.iso.org/obp/ui/#iso:std:iso-iec:27000:ed-3:v1:en

Joglekar, P., & Pise, N. (2016). Solving Cyber Security Challenges using Big Data. *International Journal of Computer Applications*, *154*(4), 9-12. https://doi.org/10.5120/ijca2016912080

Kaiafas, G., Varisteas, G., Lagraa, S., State, R., Nguyen, C. D., Ries, T., & Ourdane, M. (2018). Detecting malicious authentication events trustfully. *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, 1-6. https://doi.org/10.1109/NOMS.2018.8406295

Kritzinger, E., & Smith, E. (2008). Information security management: An information security retrieval and awareness model for industry. *Computers & Security*, *27*(5-6), 224-231. https://doi.org/10.1016/j.cose.2008.05.006

Kumar, P., Kumar, P., Zaidi, N., & Rathore, V. S. (2018). Analysis and Comparative Exploration of Elastic Search, MongoDB and Hadoop Big Data Processing. En M. Pant, K. Ray, T. K. Sharma, S. Rawat, & A. Bandyopadhyay (Eds.), *Soft Computing: Theories and Applications* (Vol. 584, pp. 605-615). Springer Singapore. https://doi.org/10.1007/978-981-10-5699-4_57

Labrinidis, A., & Jagadish, H. V. (2012). *Challenges and Opportunities with Big Data*. 2.

Ley Orgánica de Protección de Datos Personales. (2021). *Ley Orgánica de Protección de Datos Personales*.

Ley Orgánica de Telecomunicaciones. (2015). *Ley Orgánica de Telecomunicaciones*.

Liu, R., Li, Q., Li, F., Mei, L., & Lee, J. (2014, octubre). *Liu2014.pdf*. https://doi.org/10.1109/SOLI.2014.6960762

Lněnička, M., Máchová, R., & Komárková, J. (2017). Components of Big Data Analytics for Strategic Management of Enterprise Architecture. *Conference: 12th International Conference on Strategic Management and Its Support by Information Systems 2017*, 8.

Maeda, N., Agetsuma, N., Kamimura, K., Suenaga, Y., Takebayashi, S., & Yamashita, K. (2018). *Achieving Greater Work Efficiency in Systems Failure Analysis Using Elastic Stack*. *16*(2), 6.

Mujawar, S., & Kulkarni, S. (2015). Big Data: Tools and Applications. *International Journal of Computer Applications*, *115*(23), 7-11. https://doi.org/10.5120/20289-2113

Nadeem, S. F., & Huang, C.-Y. (2018). Data Visualization in Cybersecurity. *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 48-52. https://doi.org/10.1109/CSCI46756.2018.00017

P., R. M., & C., I. M. (2019). 2551-Article Text-4005-1-10-20191230.pdf. *International Journal of Advanced Science and Technology*, *Vol. 28, No. 19*, 425-432.

Pérez Marqués, M. (2015). *Big Data ecnicas herramientas y aplicaciones* (Primera Edición). Alfaomega Grupo Editor, S.A. de C. V.

Rohit, Gupta, B., Kumar, R., & Kumar, A. (2018). Towards Information Discovery On Large Scale Data: State-of-the-art. *2018 International Conference on Soft-Computing and Network Security (ICSNS)*, 1-9. https://doi.org/10.1109/ICSNS.2018.8573666

Roji, K., & Sharma, G. (2019). *Cyber Security Challenges and Big Data Analytics*. 4.

Subburaj, T., Department of Computer Applications, Kalasalingam Academy of Research and Education, Krishnankoil - 626126, Tamilnadu, India, Suthendran, K., & Department of Information Technology, Kalasalingam Academy of Research and Education, Krishnankoil - 626126, Tamilnadu, India. (2018). DigitalWatering Hole Attack DetectionUsing Sequential Pattern. *Journal of Cyber Security and Mobility*, *7*(1), 1-12. https://doi.org/10.13052/jcsm2245-1439.711

Talas, A., Pop, F., & Neagu, G. (2017). Elastic stack in action for smart cities: Making sense of big data. *2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, 469-476. https://doi.org/10.1109/ICCP.2017.8117049

Tounsi, W., & Rais, H. (2018). A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & Security*, *72*, 212-233. https://doi.org/10.1016/j.cose.2017.09.001

Veiga, A. D., & Eloff, J. H. P. (2007). An Information Security Governance Framework. *Information Systems Management*, *24*(4), 361-372. https://doi.org/10.1080/10580530701586136

wazuh. (2021a). *Overview—User manual · Wazuh 4.1 documentation*. Overview. https://documentation.wazuh.com/current/user-manual/overview.html

wazuh. (2021b). *Welcome to Wazuh · Wazuh 4.1 documentation*. https://documentation.wazuh.com/current/index.html