Pontificia Universidad Católica del Ecuador | SEDE ESMERALDAS

ESCUELA DE SISTEMAS Y COMPUTACIÓN

TESIS DE GRADO

ANÁLISIS DE ARCHIVOS LOGS DE UN SITIO WEB

PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN

AUTOR

RIVAS GRACIA BRYAN ISAAC

ASESOR

MGT. XAVIER QUIÑÓNEZ KU

Esmeraldas – Abril, 2019

Tesis de grado aprobada luego de haber dado cumplimiento a los requisitos exigidos, previo a la obtención del título de INGENIERO EN SISTEMAS Y COMPUTACIÓN.

TRIBUNAL DE GRADUACIÓN

Título: "ANÁLISIS DE ARCHIVOS LOGS DE UN SITIO WEB"

Autor: RIVAS GRACIA BRYAN ISAAC

Mgt. Xavier Quiñónez Ku f.-……
Asesor/a

Mgt. José Luis Carvajal f.-…….
Lector #1

Mgt. Juan Casierra Cavada f.-…….
Lector #2

Mgt. Xavier Quiñónez Ku f.-……
Director de Escuela

Esmeraldas, Abril del 2019

AUTORÍA

Yo, **BRYAN ISAAC RIVAS GRACIA** portador de la cédula de identidad No. **080326207-0**, declaro que los resultados obtenidos en la presente investigación como tesis de grado son absolutamente originales, auténticos y personales.

En tal virtud, declaro que el contenido que se desprenden de éste documento son y serán de mi exclusiva responsabilidad legal y académica.

BRYAN ISAAC RIVAS GRACIA

CI 080326207-0

CERTIFICACIÓN

Mgt. Xavier Quiñónez Ku, docente investigador de la PUCE-Esmeraldas, certifica que:

La tesis de grado realizada por BRYAN ISAAC RIVAS GRACIA, bajo el título "ANÁLISIS DE ARCHIVOS LOGS DE UN SITIO WEB", reúne los requisitos de calidad, originalidad y presentación exigibles a una investigación científica y que han sido incorporadas al documento final las sugerencias realizadas, en consecuencia, está en condiciones de ser sometida a la valoración del Tribunal encargado de juzgarla.

Y para que conste a los efectos oportunos, firma la presente en Esmeraldas, abril del 2019

Mgt. Xavier Quiñónez Ku

Asesor

AGRADECIMIENTO

Agradezco primeramente a Dios por darme una familia que siempre estuvo pendiente de mí y apoyándome para poder cumplir mi meta de ser un profesional.

A mi madre, Juana María Gracia Loor, por motivarme y apoyarme para poder salir adelante de forma incondicional.

A mi padre, Walter Fernando Rivas Rangel, por que a pesar de las circunstancias siempre me apoyó para poder alcanzar mi meta y convertirme en un verdadero profesional.

A Jhormy Wilson González Perlaza, quien se convirtió en un segundo padre para mi; siempre me aconsejaba y me apoyaba para que no tirara la toalla en el proceso de mi formación académica, y ahora profesional.

A mi hermana, Dayra Fernanda Rivas Gracia, que a pesar de las diferencias que teníamos como hermanos, siempre ha estado conmigo apoyándome en todo momento.

Y al resto de mi familia, que en alguna ocasión me ha brindado un consejo de motivacion para que no decaiga y siga luchando por mis sueños, éste logro es para todos ustedes.

DEDICATORIA

El trabajo de investigación está dedicado a mi familia, quienes siempre me han apoyado en el transcurso de mi carrera universitaria; a mi padre, madre, padrastro y hermana, quienes siempre han estado ahí, sirviéndome de motivación para poder seguir adelante.

Índice General

Índices Ilustraciones

Índices Tablas

RESUMEN

La presente investigación se llevó a cabo con la finalidad de monitorear y analizar los archivos logs que se generan el en servidor que aloja el sitio web de la PUCE Sede Esmeraldas, con el objetivo de prevenir y alertar los diferentes ataques, vulnerabilidades y problemas que pueden originar.

Para llevar a cabo la investigación se realizó una comparación de varias herramientas analizadores de logs, pertenecientes a software libre y que brinde una gran flexibilidad y potencia para su uso. De éstas se concluyó que la herramienta que cumple con éstos requisitos es la llamada ELK, la cual se encuentra conformada por 3 herramientas denominadas: Elasticsearch, Logstash y Kibana; dónde Elasticsearch funciona como la BD de la aplicación, Logstash es el recolector que almacena todos los archivos logs generados en el servidor y Kibana es la interfaz gráfica qué brinda una representación gráfica en forma de gráficos estadísticos de los diferentes atributos que se quieren examinar.

Al analizar los archivos logs del servidor antes mencionado, se obtuvo como resultado que al sitio web se accede desde ubicaciones, que tal vez ni la propia Institución se imaginaría; además de que se logró detectar qué, cierta cantidad de accesos que se realizaron son originados por bots; también permite conocer los diferentes códigos HTTP que se originan el servidor al momento de realizar una determinada acción y a qué ubicación se está tratando de acceder, etc.

De ésta forma se puede concluir que la utilización de la herramienta es de gran ayuda al momento de monitorear y analizar los archivos logs, esto se debe a que permitió localizar la ubicación de los diferentes equipos clientes que se encuentran accediendo al sitio web, también permitió conocer los diferentes sistemas operativos que se hace uso para acceder a éste y a su vez de sistemas extraños que son pertenecientes a bots, etc.

Palabras clave: análisis**,** archivos logs, sitio web

ABSTRACT

The present investigation was carried out with the purpose of monitoring and analyzing the log files generated by the server that hosts the website of the PUCE Sede Esmeraldas, with the objective of preventing and alerting the different attacks, vulnerabilities and problems that may originate.

To carry out the research, a comparison was made of several log analyzer tools, belonging to free software and that offer great flexibility and power for its use. From these it was concluded that the tool that meets these requirements is the so-called ELK, which is made up of 3 tools called: Elasticsearch, Logstash and Kibana; where Elasticsearch works as the BD of the application, Logstash is the collector that stores all the logs generated on the server and Kibana is the graphical interface that provides a graphical representation in the form of statistical graphs of the different attributes that we want to examine.

When analyzing the logs files of the aforementioned server, it was obtained as a result that the website is accessed from locations, which perhaps even the Institution could not imagine; besides that it was possible to detect what, a certain amount of accesses that were made are originated by bots; it also allows to know the different HTTP codes that originate the server when performing a certain action and to what location it is trying to access, etc.

In this way we can conclude that the use of the tool is of great help when monitoring and analyzing the logs, this is because it allowed locating the location of the different client computers that are accessing the website, it also allowed know the different operating systems that are used to access this and in turn strange systems that are belonging to bots, etc.

Keywords: analysis, log files, website

Introducción

Presentación de la investigación

La presente investigación consiste en resaltar la importancia que brinda el uso de los archivos logs, ya que éstos tipos de archivos almacenan información relevante de las acciones o eventos que se realizan en un software.

Actualmente la tecnología a evolucionado en gran magnitud, a tal punto que las personas ya no necesitan ir a un lugar para realizar una determinada compra o venta de un producto o servicio, sino qué por medio de un sitio web se pueden realizar todas éstas tareas y desde la comodidad del hogar.

Pero así como la tecnología ha brindado beneficios a la sociedad, también la ha dejado expuesta a ciertos peligros como el robo de información, suplantación de identidad, etc; esto se debe a que para poder usar éstos tipos de sistemas, el usuario o cliente debe proporcionar cierta información personal y confidencial, y si no se lleva cabo un buen control y gestión de seguridad en la aplicación, el sistema podría quedar expuesto a una gran variedad de ataques.

Con el correcto uso de los archivos logs y la información que éstos almacenan, se puede logar prevenir los problemas mencionados anteriomente, debido a que guardan información relevante de los clientes que acceden al sistema como es la ubicación del cliente, a qué parte de la página está accediendo, reconoce el sistema operativo que se está haciendo uso, entre otros datos importantes; permitiendo llevar una correcta gestión de seguridad y monitoreo del sitio web que se está analizando.

Planteamiento del problema

Con el paso del tiempo la tecnología ha ido evolucionando a pasos gigantescos facilitando el uso, de cierta forma, la realización de múltiples habilidades y tareas que comúnmente las personas realizan en su vida cotidiana, dónde ya no es necesario ir a un lugar específico para conseguir algún producto o servicio, sino que por medio de un sitio web las organizaciones ofrecen sus servicios en línea y el cliente con un clic es capaz de realizar alguna transferencia o compra de un producto deseado.

Una de las ciencias que más ha avanzado en los últimos años es la Informática, la cual ha influido en el desarrollo de la red de redes, Internet, para lograr un mundo

> interconectado, donde la comunicación y acceso a los recursos es mucho más fácil. Las aplicaciones web han facilitado dicho proceso, mediante su facilidad de uso y presencia en redes sociales, buscadores, sitios de comercio, información interactiva, las cuales han sido ampliamente adoptadas dentro de las organizaciones para soportar las funciones claves del negocio. (Hernández y Porven, 2016)

Ésta nueva implementación de la tecnología con los negocios ha ocasionado un gran impacto tanto personal como profesional, ya que actualmente el usuario para poder realizar una determinada acción debe confiar y brindar cierta información personal a aquella organización como es el nombre, número de cedula, tarjetas de créditos, dirección donde vive, etc.

A causa de lo ya mencionado anteriormente, los servidores web se han convertido en uno de los blancos preferidos por los diferentes atacantes o hackers, debido a que éstos tienen el conocimiento del tipo de información que se almacena y de las cosas que se podría realizar, ocasionando problemas tales como robo de información, suplantación de identidad, ente otros delitos informáticos.

Hernández y Mejia (2015) manifiestan: “En la actualidad el riesgo para los sistemas informáticos ha aumentado debido a un crecimiento en la complejidad en las tecnologías de la información.”.

Los avances tecnológicos, a pesar de brindar un gran aporte a la sociedad, se encuentran expuestos a una gran variedad de amenazas, y si no se posee una buena administración de los diferentes sistemas, la empresa queda expuesta a múltiples factores como: alteración de la información, suplantación de identidad de algún usuario, robo de la información, robo económico, etc. Ocasionando una cadena de consecuencias como: pérdidas materiales, inconsistencia de información, pérdidas económicas, desprestigio de la empresa, quiebre de la empresa y, tal vez, un fuerte juicio legal.

Mantener un monitoreo constante de los diferentes logs que se generan, con el fin de proceder a analizar los eventos detectados en el equipo servidor, es muy importante; ya que permite actuar de forma oportuna ante cualquier dificultad o percance que se

presente, permitiendo evitar la manipulación o pérdida de la información que maneja la institución de forma privada.

Avella, Calderón y Mateus (2015) afirman: “Las herramientas y aplicaciones tecnológicas, generan registros de seguridad que normalmente son almacenados localmente y contienen información acerca de los estados, comportamientos y cambios que ocurren en cada dispositivo”. Sin embargo, analizar los archivos que se generen de cada equipo por separado no es suficiente y es deficiente, ya que en ciertos casos la actividad de un equipo se encuentra directamente relacionada con la de otro equipo, por lo tanto, si se llegase a analizar cierta actividad no tendría ningún sentido porque faltaría información para completar su análisis.

La revisión y análisis de éstos archivos es sumamente compleja y tediosa, y más cuando se genera un sinnúmero de archivos diariamente. Para darle solución a éste problema es conveniente almacenar todos los archivos generados en un solo equipo y desde éste contenedor comenzar el análisis.

Justificación

La presente investigación sobre los archivos logs consiste en resaltar la gran importancia que estos aportan en la administración de un servidor que contiene un sitio web, ya sea para una empresa u organización; dando a conocer que estos tipos de archivos no solo son archivos comunes y que no solo se encuentran en un determinado sistema o programa, sino que están en la gran mayoría de sistemas que comúnmente se hace uso y que la mayoría de usuarios no posee conocimiento de su existencia.

Al momento de hacer uso de un servidor web, es necesario llevar la correcta administración y monitoreo de éstos equipos, ya que se realizan una gran cantidad de actividades relevantes en la red y mediante el correcto análisis de los archivos logs generados se puede llegar a conocer información que resulta muy útil para el debido control, como podría ser la información de los usuarios que acceden al sitio o servicio alojados, si dio inicio o accedió alguna aplicación, el Sistema Operativo que está

haciendo uso, el tiempo en que realizó una determinada acción, el rendimiento del servidor, el origen de algún error y de cómo es su comportamiento en la red.

Debido a que el acceso hacia éstos equipos se encuentra de forma pública, es decir cualquier persona puede acceder a éstos, se encuentra expuesto a múltiples ataques; de esta forma se podrá hacer énfasis en el uso de estos tipos de archivos, ya que la información que se maneja en este tipo de sistema es privada y por ende es necesario saber cómo protegerla

Objetivos

General

Analizar los diferentes archivos logs generados mediante el testeo de diferentes herramientas que existen en la actualidad, para resaltar la importancia que estos poseen en la administración de los sitios web de la PUCE Sede Esmeraldas.

Especifico

- Recopilar información referente al uso de archivos logs para comprender la información que estos almacenan y la aplicabilidad de los mismos.
- Evaluar diferentes herramientas dedicadas al análisis de archivos logs con el fin de seleccionar la mejor de ellas.
- Implementar la herramienta seleccionada en un entorno de desarrollo en el Departamento de TIC.
- Examinar los archivos logs generados en la herramienta seleccionada.

CAPÍTULO I

1. Marco teórico

1.1. Antecedentes

Los servidores web son ordenadores, de cierta forma, especiales que realizan una función muy aparte de un ordenador normal, donde atienden las diferentes peticiones que se realizan en otros ordenadores manejados por usuarios o clientes y a la vez le facilita la información solicitada en el mismo.

Cuando un ordenador accede al sitio web, el servidor registra cada visita que recibe de forma automática la cual posee la fecha y hora, un identificador (ID) único, la acción realizada, la dirección IP del cliente, el navegador, el Sistema Operativo usado, etc. Permitiendo llevar un mejor control del sitio web establecido.

> Los mensajes de los logs son muy importantes para los administradores de sistemas para entender el comportamiento previo del sistema. Pueden mostrar pistas de lo que está pasando en el sistema, en las redes, así como recolectar datos sobre el rendimiento y sobre todo la detección de intrusiones. (Mellouk, 2016)

La gran cantidad de información que presenta los archivos logs son esenciales para la administración de un sistema o de un sitio web, ya que los servicios o productos que se brinden en estos sitios son los ingresos de una empresa o institución y por ende, en el momento que uno de los sistemas falle representa una gran pérdida tanto social como económica, ya que se pierden ingresos y la reputación de la empresa se vería severamente afectada; por lo cual la intervención de los archivos logs resultan necesarios ya que el correcto análisis de estos, permiten verificar el rendimiento del equipo y prevenir futuros fallos.

> Las instituciones financieras, al igual que muchas otras empresas en diferentes sectores, han tomado conciencia a través de los años de los riesgos e incertidumbres que surgen de las fallas en sus operaciones y recientemente, de manera particular, de las fallas en los sistemas de información y la

> infraestructura tecnológica, paradigmas de suma importancia en la actualidad. De tal manera que los fraudes, la interrupción de la actividad operativa y la responsabilidad legal se han convertido en una amenaza constante para cualquier empresa. (Dávila, Ortiz, y Cruz, 2016)

Cuando se habla de instituciones financieras la probabilidad de que ocurra un fallo en sistema debería ser minina, ya que en aquella empresa se maneja el dinero, por lo cual la pérdida de un porcentaje del dinero que tenga guardado uno de sus clientes es un gran fallo para la empresa, lo que podría abarcar múltiples problemas legales, pérdidas de ingresos, perdida de clientes por la inseguridad que posee, entre muchas otras cosas. Los archivos logs dentro de este tipo de instituciones es lo primordial, ya que como son instituciones que manejan dinero están sumamente propensas a ataques de hackers realizando: clonación de los sitios, correos falsos, continuo ataque a los servidores de forma directa, etc.; y así poder hacerse de las suyas con aquel bien.

> La informática ha traído grandes beneficios a la humanidad, pero también ha permitido la aparición de nuevos peligros relacionados con la seguridad de la información. Esta situación es un fenómeno de impacto mundial, y se evidencia en la utilización de la información con fines delictivos. (Gómez, Candela, y Sepúlveda, 2013)

Debido a la gran importancia y valor que posee la información, obtenerla se ha convertido en unos de los propósitos claves de los hackers; pero con el correcto análisis de los archivos logs se podría anticipar los posibles ataques que se harían por fuentes externas, además de conocer esos pequeños huecos dentro del sistema, por los cuales se podrían dar los ataques, además de la verificación de la seguridad del sitio y resguardando los datos de los usuarios que confían en aquella institución para así ofrecer un óptimo servicio.

Dávila, Ortiz, y Cruz (2016) manifiestan: “Considerando que los procesos y sistemas son desarrollados y administrados por personas, son ellas quienes causan los eventos de riesgo de operaciones al hacer algo que no deberían haber hecho, o no hacer algo que debieron hacer”. Las fallas en los sistemas de las instituciones muchas veces son dadas por la falta de conocimiento y de experiencia de quienes los desarrollan, los cuales dejan pasar muchas cosas por alto, como es la seguridad, y solo se enfocan en el funcionamiento; dejando a los sistemas expuesto a riesgos con los cuales personas

sin mucho conocimiento en hacker rompen fácilmente la seguridad de los sitios web, haciendo de las suyas con la información que se brindan y se almacenan en estos.

Cuanto a desarrollar un sistema se habla, se debe tratar con personas que ya tengan un amplio conocimiento en este ámbito y antes de lanzarlo a producción, realizarles una infinidad de pruebas para verificar la calidad que aparenta brindar, en la actualidad existen una gran cantidad de empresas que se dedican a este tipo de negocios, los cuales otorgan todos posibles fallos que el sistema tendría para así seguir a su mejoramiento.

Según Ferreira (2015) afirma: “Los logs son relevantes para la identificación de características, demandas y necesidades de los usuarios”. Los archivos logs además de aportan una gran información para la seguridad del sitio web, también se puede utilizar la misma información como opción para la toma de decisiones del sitio web, ya que brinda la información de las pagina más visitada por los usuarios, montándolo en un sitio que sea de venta de productos, se podrá obtener los productos que más visitan y necesiten los usuarios pudiendo resultar estos archivos una gran ayuda no solo para la seguridad del sitio, sino también como una gran ventaja de marketing para beneficio de la empresa y obtener mayores ingresos.

1.2. Base teórica

1.2.1. Servidores web

A mediados de la los 90, la aparición del internet ha tenido un gran impacto y se ha extendido de una forma jamás pensada, permitiendo la integración entre sistemas tanto hardware como software. Desde entonces gran variedad de empresas ha ido mejorando cada vez más la tecnología para tener una integración más óptima con los sistemas. Éstas empresas se dieron cuenta de que sería imposible crear una plataforma que se integre de forma individual, por lo cual para acabar con éste problema se buscó un lenguaje de intercambio de información abarcado con el standard que existen hoy en día, surgiendo de éste modo los Servicios Web basados en XML.

Cuando se habla de un servidor web, se hace referencia aquel equipo que suministra servicios a los usuarios o terminales, a través de un sitio o aplicación web, los cuales acceden por medio de un navegador mediante el protocolo HTTP perteneciente a la capa de modelo OSI.

1.2.1.1. Funcionamiento

Básicamente, la función de un servidor web consiste en el almacenamiento de los diferentes archivos de un sitio web y a la vez emitirlos por la red para que los usuarios puedan acceder a estos. El servidor se ejecuta en un ordenador, en donde se encuentra a la espera de las peticiones que los clientes o usuarios realicen por medio de un navegador y a su vez responde a estas peticiones por medio de una página que se mostrará en el respectivo navegador en donde se realiza la petición.

Cada servidor que está conectado a la red posee una dirección IP única, la cual permite identificar del resto de equipo de la red, además de que por medio de ésta se accede al equipo, incluyendo los dispositivos móviles y los ordenadores. Cuando lo que se quiere es ver o navegar un sitio web, se envía un pedido hacia la dirección IP del servidor que aloja los archivos objetivos, el servidor web resuelve éste pedido y devuelve los datos hacia la dirección IP del solicitante.

1.2.1.2. Arquitectura

Cuando se habla de arquitectura web, se hace referencia a toda aquella tecnología que se hace uso para poner en funcionamiento el servidor web, permitiendo a los diferentes usuarios visualizar el contenido ofrecido o subido a la red.

> La arquitectura web se refiere a la programación de una aplicación web, lo cual incluye tener un servidor operativo (Apache, por ejemplo) y una base de datos (en MySQL o cualquier otro lenguaje de base de datos con el cual se disponga de conector). El núcleo de la aplicación se desarrollará, básicamente, en un lenguaje como PHP o Java (mediante JSP), estando acompañado por código HTML y por JavaScript. (Granados, 2014)

Ilustración 1 Estructura básica de una aplicación web

Fuente: (Granados, 2014)

La estructura de un servidor web consiste en:

1. El usuario realiza una petición al servidor que contiene el sitio web o aplicación web por medio un cliente llamado navegador.
2. El servidor posee una conexión hacia una base de datos que es la que posee toda la información para el funcionamiento del sitio web.
3. El servidor envía la respuesta al cliente después de obtener o modificar la información de la respectiva base de datos.
4. Por último, el cliente obtiene esta información enviada desde el servidor, la cual la presenta al usuario de forma entendible.

1.2.1.2.1. Modelo de capas

El modelo OSI es un punto a destacar dentro de la arquitectura de una red, ya que esta describe como se da el proceso de transmisión de datos. Dentro de este modelo existen 7 capas, de las cuales el usuario solo interactúa con dos de estas, la primera (capa física) y la última (capa de aplicación).

De la Fuente (2012) manifiesta: “Este marco de trabajo estructurado en capas, aun siendo puramente conceptual, puede utilizarse para describir y explicar el conjunto de protocolos reales que, como veremos, se utilizan para la conexión de sistemas.”

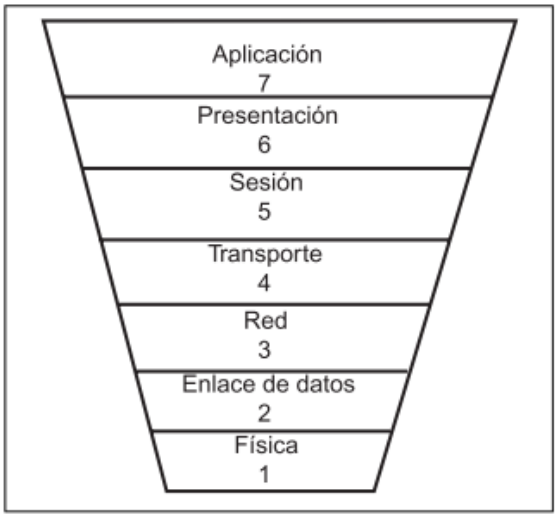


Ilustración 2 Estructura de las capas de modelo OSI

Fuente: (De la Fuente, 2012)

La imagen representa la estructura de las diferentes capas que conforma el modelo OSI de arriba hacia abajo. Una forma de ilustrar de mejor manera esta estructura de este modelo es con la pirámide invertida.

Dentro de las 7 capas de Modelo OSI se tiene:

1. **Capa Física**

Las tramas que se originan en la capa de enlace de datos se transforma en una secuencia única de bits, los cuales propagan por medio del entorno físico de red. También especifica los diferentes aspectos físicos (características materiales, eléctricas, etc.) de como la interfaz de internet del ordenador está conectada al

cableado. En lo que respecta el ordenador receptor, recibe una secuencia de bits (puede ser 0 o 1), la cual se encarga la capa física.

2. Capa de Enlace de Datos

Los paquetes que alcanzan esta capa se ubican en unidades de datos (tramas), estos se definen por la arquitectura de la red que se utiliza. Esa capa se encarga de gestionar la transferencia de datos por medio del enlace físico de comunicación hasta que llegue al ordenador receptor, identificando cada ordenador participante que se encuentra en la red de acuerdo a la dirección de hardware.

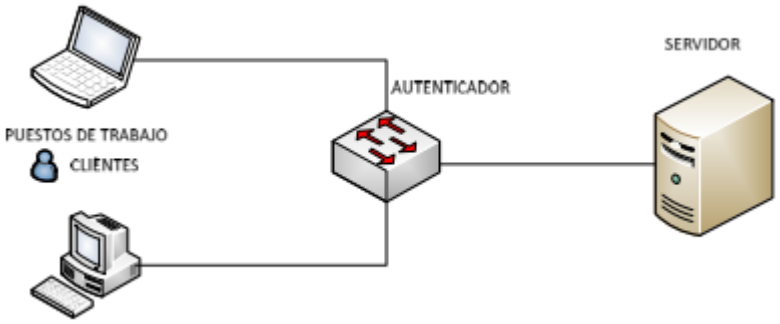


Ilustración 3 Capa de enlace de datos

Otra de las funciones que realiza es encargarse de que las tramas lleguen a su ordenador receptor sin ningún error, por lo que los protocolos que se hace uso en esta capa adjunta al final de cada trama un CRC (Chequeo de Redundancia Ciclica o Cyclical Redundancy Check). El CRC es un valor calculado en el participante emisor y receptor, en donde si en estos dos ordenadores coincide el mismo valor, significa que el envío y recepción de trama no sufrió ningún fallo y llego de forma íntegra.

3. Capa de Red

Se encarga de dirigir los paquetes, se puede decir que, determina la ruta que los diferentes paquetes deben seguir. En esta capa las direcciones lógicas (dirección IP de un ordenador conectado a una red) se convierten en una dirección física (dirección de hardware, tarjeta de interfaz de red del respectivo ordenador).

Los routers son los que se encargan de ejercer esta acción, por medio de protocolos de encapamiento se determina las rutas que deben seguir los diferentes paquetes que se transmiten.

La capa de red posee dos formas de funcionar:

- **Mediante datagramas. -** Donde los diferentes paquetes de datos son independientes, sin necesidad de que haya previamente una comunicación.
- **Mediante circuitos virtuales. -** Se establece la conexión entre los diferentes ordenadores a comunicar, por lo que forma un circuito donde los paquetes podrán circular.

Además, también ofrece dos tipos de servicios:

- **Servicios orientados. -** Donde el primer paquete contiene sus direcciones de destino, la cual establece una determinada ruta de los paquetes que tengan la misma conexión.
- **Servicios no orientados. -** Donde cada paquete contiene su dirección destino, los cuales eligen una ruta orientada por la técnica de encapamiento.

4. Capa de Transporte

Controla el flujo de los paquetes enviados en la comunicación de los diferentes ordenadores palpitantes, ya que estos datos no solo deben ser entregados sin errores, sino también en la secuencia en la que se envía. Otra funcionalidad que realiza esta capa es la de conocer el tamaño de los paquetes, por lo que estos deben tener un tamaño requerido para las capas inferiores.

La comunicación se realiza por ordenadores participantes del mismo nivel, por lo que el número de paquetes enviados por el participante emisor debe ser acordado o aceptado con el número de paquetes que recibirá el participante receptor.

Esta capa resulta muy útil cuando el ordenador emisor envía demasiados paquetes al ordenador receptor, en donde el receptor solo toma aquellos paquetes que alcance a aceptar enviando una "señal de ocupado", y cuando este ya haya procesado los paquetes y se encuentre en estado de listo para recibir más envía una señal de "listo" para recibir los paquetes restantes.

Existen dos protocolos dependiendo de la orientación a conexión:

- **UDP (Protocolo no orientado a conexión). -** Consiste en que cada datagrama contiene información útil para poder llegar a la dirección destino, pero existe una verificación de que haya llegado a su destino, ni que error hubo; además

de que los paquetes podrían llegar en cualquier orden y no en la forma de que fueron enviados.

- **TCP (Protocolo orientado a conexión).** - Los paquetes ya poseen un orden y un control de errores; además de permitir la distinción de las diferentes aplicaciones que trabajan en un mismo ordenador por medio de puertos.

5. Capa de Sesión

Se encarga de establecer la respectiva conexión o enlace de comunicación entre el ordenador, emisor tanto como el receptor.

Ilustración 4 Capa de sesión

Fuente: (De la Fuente, 2012)

Cuando se establece la conexión entre los ordenadores participantes, esta capa comienza hacerse cargo de posicionar los diferentes puntos de control para la secuencia de datos, proporcionando cierta tolerancia cuando se produzca un evento de fallo de comunicación. En caso de que la comunicación llegara a fallar entre los ordenadores participantes, cuando se restablezca solo se deberá enviar los paquetes situados en el punto de control anterior al del fallo y se evitaría enviar de nuevo todos los paquetes incluida la sesión.

Para establecer una conexión con el enfoque de que los datos vayan del emisor al receptor existen dos protocolos que proporciona la capa de sesión: la comunicación sin conexión y la comunicación orientada a la conexión.

Los protocolos orientados a la conexión consisten en que los ordenadores que se encuentran conectados trabajan de forma cooperativa, donde establecen una concordancia sobre los parámetros relativos para la ubicación de los puntos de

control, manteniendo una comunicación durante la trasferencia de estos para después terminar de manera simultánea la sesión.

Los protocolos de comunicación sin conexión tienen similitud a un sistema de correo, donde existen varias direcciones que son pertinentes para el envió de datos. Estos datos llegarán a las direcciones destino sin necesidad de que el ordenador receptor otorgue un permiso previo.

Los servicios que ofrece la capa de sesión son:

- **Control de sesión. –** Intervienen el emisor, receptor y la sesión establecida en sí.
- **Control de concurrencia.** - Evita que dos o más operaciones se ejecuten al mismo tiempo en el mismo equipo.
- **Mantenimiento de puntos de reanudación.** - Reanuda la conexión cuando la comunicación falle.

6. Capa de Presentación

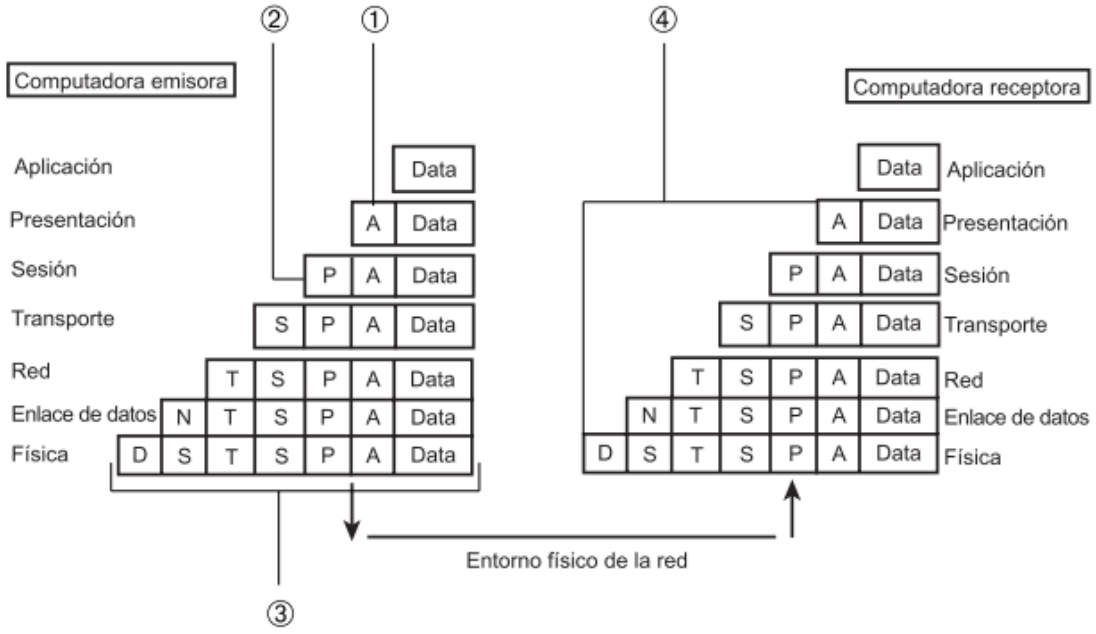
Se la puede considerar como el traductor del Modelo OSI, se encarga de obtener los datos que se transmiten por medio de la red de la capa de aplicación y los transforma a un formato que sea entendible para los ordenadores.

Entro otras aplicaciones también se encarga del cifrado de datos, siempre y cuando la aplicación usada requiera el cifrado, además disminuir el tamaño por medio de la compresión de los mismo.

7. Capa de Aplicación

Esta capa, además de proporcionar acceso general a la red, se encarga de ofrecer la interfaz y servicios de las aplicaciones de usuarios; también proporciona los diferentes servicios de red que se encuentran relacionados con las aplicaciones de usuarios como son: envió de archivos, consulta a las bases de datos y administración de mensajes. Básicamente, cada programa de aplicación que el usuario tenga en su computadora es suministrado para la capa de aplicación.

En la ilustración 5 se muestra como los datos, de la computadora emisora, bajan por la pila OSI y suben por otra pila OSI, de la computadora receptora.



1. Encabezado de la capa de aplicación.
2. Encabezado de la capa de presentación.
3. Paquete con todos los encabezados de las capas OSI.
4. Los encabezados se van suprimiendo a medida que los datos suben por la capa OSI.

Ilustración 5 Capa de aplicación

Fuente: (De la Fuente, 2012)

1.2.1.2.2. Plataformas para el desarrollo en las capas servidor

En la actualidad existen muchas plataformas para el desarrollo en el lado del servidor, de las cuales 3 son las más utilizadas:

- **PHP (Personal Homes Pages)**

  Es un popular leguaje de programación que por lo general se presenta en forma de scripts, el cual se encuentra incrustado o emparejado con lo que sería HTML. Palomo (2013) manifiesta: "Es un lenguaje interpretado con una sintaxis similar a la de C++ o JAVA. Aunque el lenguaje se puede usar para realizar cualquier tipo de programa, es en la generación dinámica de páginas web donde ha alcanzado su máxima popularidad."

Además, es un lenguaje de código abierto, el cual aporta una gran flexibilidad y facilidad al momento de comenzar el desarrollo de alguna aplicación y también se basa en la programación orientada a objetos.

- **ASP.NET**

Es una plataforma originaria o propia de la de empresa de Microsoft, la cual hace uso del de un framework muy utilizado como es .NET.

> ASP.NET es el nombre con el que se conoce la parte de la plataforma .NET que permite el desarrollo y ejecución tanto de aplicaciones web como de servicios web. En ASP.NET, no obstante, las aplicaciones web se suelen desarrollar utilizando formularios web, que están diseñados para hacer la creación de aplicaciones web tan sencilla como la programación en Visual Basic. (Berzal, Cortijo, y Cubero, 2010)

También al ser propio de Microsoft, posee una gran integración con las diferentes aplicaciones de escritorios propias de la misma empresa. Una de las ventajas de esta plataforma es que, para poder apreciar el desarrollo de las aplicaciones, primero se debe compilar el condigo antes de su utilización, permitiendo que la ejecución en el servidor sea mucho más rápida.

- **JSP**

Se trata de una plataforma que fue desarrollada por Sun Microsystems y que actualmente se encuentra en propiedad de Oracle.

> La tecnología JavaServer Pages permite la separación del contenido estático en una plantilla del contenido dinámico que se inserta en la plantilla estática. Esto simplifica enormemente la creación de contenido. La separación está respaldada por beans diseñados específicamente para la interacción con objetos del lado del servidor y por el mecanismo de extensión de etiquetas. (Chung, 2013)

Al igual que ASP, en el servidor de debe compilar y ejecutar para facilitar el acceso a los diferentes clientes. Además, incluye varias tecnologías de Java como son los servlets.

1.2.1.2.3. Herramientas de desarrollo orientadas a servidor de aplicaciones web

Para poder tener un servidor web funcionando se hace necesario la implementación de algunas herramientas y componentes, dentro de este tipo de herramienta se cuenta con una infraestructura de red, que básicamente se refiere a todo el hardware de un ordenador, el cual se encuentra configurada de forma singular o especial, en comparación con un ordenador normal y que tiene una conexión a internet.

Además de este, contiene un servidor web, el cual comprende la parte interna del ordenador o más conocido como el software; este es instalable y configurable con el fin de obtener un servidor operativo y robusto. Existen varios tipos de software que permiten a un ordenar dar la función de un servidor, dentro es estos se tiene los más populares:

- **Apache**

Originario de Linux, aunque existen varias versiones para el SO Windows, por lo que es un SO gratuito (código abierto) y multiplataforma. Labrador (2010) indica: "Linux es un sistema operativo de la familia Unix creado mediante la política de "código abierto". Estas características implican un gran ahorro en los costes de instalación de los equipos, pero también una mayor especialización por parte del personal informático".

Tiene una configuración por defecto que brinda una gran experiencia al momento de trabajar en un área local, pero al momento de ya llegar a un nivel más avanzada se necesita de algunos permisos para poder editar ciertas carpetas especiales o archivos que se encuentran dentro del directorio del servidor.

- **Internet Information Server (IIS)**

Servidor exclusivo para Windows, a diferencia del anterior hablado, tiene restringidas ciertas opciones; además de que no es de código abierto.

> El rol de Servidor web (IIS) en Windows Server 2012 R2 proporciona una plataforma modular y extensible para hospedar sitios web, servicios y aplicaciones de manera confiable. IIS permite a las empresas compartir información con usuarios en Internet, una intranet corporativa o una

> extranet de una compañía. IIS es una plataforma web unificada segura y fácil de administrar que integra IIS, ASP.NET, servicios FTP, PHP y Windows Communication Foundation (WCF). (Tulloch, 2013)

Para poder configurar el sitio web alojado dentro del servidor IIS, se debe acceder por medio de un panel que proporciona el mismo servidor.

Cuando se habla de servidor web, y que este contiene un sitio web o una aplicación, se habla de contiene o debe contener si o si una Base de Datos para su correcto funcionamiento. Rubinos y Nuevo (2011) afirman: "Se define una base de datos como una serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular".

Al igual que los lenguajes, también existen una gran variedad de BD en el mercado; aunque una de las BD más estables y usadas es MySQL, además de al momento de instalar con el servidor Apache, brinda una herramienta llamada PhpMyAdmin, permitiendo gestionar o usar la Base de Datos de forma más sencilla y sin necesidad de ingresar por medio de la consola. También es una herramienta accesible desde un navegador, brindando una interfaz gráfica amigable y a la vez una manipulación de datos flexible de forma remota.

El sitio web es la una de los componentes más importante al momento de hablar de servidores web, este estará desarrollada por una de las plataformas anteriormente mencionadas, o por una combinación de las dos. Según Dávila, Galvis, y Vivas (2015) afirman: "El sitio web es una herramienta por medio de la cual se apoyan los procesos de enseñanza – aprendizaje; permite brindar información relevante y plantear actividades que se desarrollan combinando la enseñanza presencial con la tecnología no presencial".

Posee una infinidad de uso, como brindar información de instituciones o empresa, compra y venta de los diferentes productos y servicios que se encuentran actualmente en el mercado. Además, un uso que actualmente se les da a estos sitios es el medio de comunicación, permitiendo compartir con varios usuarios un sinnúmero de archivos multimedia a la ves de interactuar con los mismos en tiempo real.

1.2.2. Seguridad

Actualmente son muchas las personas que hacen uso de un navegador para la búsqueda de información, mirar la prensa, realizar transacciones, compras y pagos en línea por medio del internet, este tipo de acciones se han convertido en algo muy habitual en la vida diaria de muchas personas.

Gómez, Candela, y Sepúlveda (2013) afirman: "La seguridad en los servidores HTTP constituye un problema en aumento, y no todas las compañías que tienen publicada su información o servicios Web disponibles para los usuarios toman en serio esta problemática".

Los servidores web se han convertido en una herramienta muy necesaria para empresas pequeñas, medianas y grandes; donde cualquier organización, sin importar el tamaño de la misma, posee una página web para darse a conocer a la sociedad y ofertar por medio del internet los servicios o productos que esta presta, ya sea desde la página más estática que no incluya ningún contenido, hasta la más completa que brinde realizar operaciones por medio de ésta, brindando un servicio de forma directa a los usuarios.

Debido a todas estas opciones que brinda estas plataformas web, se han convertido en un blanco para la mayoría de atacantes, por consecuencia a la gran dependencia que los usuarios han llegado adoptar, donde los ataques a los diferentes servidores se han convertido en uno de los más llamativos. Por ende, cuando se procede a la creación de un sitio web el servidor debe estar protegido ante cualquier tipo de amenaza.

> Hoy en día la seguridad informática es cada vez más reconocida e importante a nivel mundial, tanto en las empresas privadas como en las instituciones públicas, ya que Internet es la principal fuente de acceso de cualquier tipo de información y por ello se han diseñado e implementado un sinnúmero de aplicaciones para casi todas las áreas del conocimiento. (Guamán, 2011)

La integración del internet en todas las áreas que existen en la actualidad, ha permitido que la mayoría de procesos que habitualmente se realizaban de manera manual se automaticen, logrando que estos procesos se desarrollen de una forma más

ágil y simple. Esta dependencia que las personas han adoptado podría ser un gran problema, debido a los diferentes fraudes que existen acerca de la seguridad.

Definir la seguridad informática, no es tan sencillo, ya que es un concepto amplio; se la puede definir como la toma de las medidas que permitan impedir la realización de diferentes operaciones que no se encuentren autorizadas sobre un determinado sistema o red, cuyas acciones o procesos podrían ocasionar daños y perdida de la información.

1.2.2.1. Principios básicos

Cuando se seguridad de habla, se hace referencia a tres aspectos fundamentales:

- **Confidencialidad. -** Consiste en que los archivos o recursos de sistema solo pueden ser accedidos por el personal autorizados.
- **Integridad.** - La información solo puede ser manipulada por el personal autorizado y de forma controlada.
- **Disponibilidad.** - La información deben estar siempre accesibles al personal autorizado.

Para obtener una seguridad completa se debe conocer ciertos aspectos como:

- ¿Qué se protege?
- ¿Contra quién se protege?
- ¿En quién se puede confiar y en quién no?
- ¿Qué tipo de empleados tienen acceso a qué activo? y ¿Por qué?
- ¿Qué tipo de acceso tendrá cierta persona de la organización?
- ¿Qué tipo de acceso tendrán los posibles clientes?

Uno de los puntos que se debe tener claro es que no hay q centrarse de que los atacantes solo se pueden originar desde el internet, sino que también puede existir dentro de la organización misma.

1.2.2.2. Seguridad en entornos web

Hernández y Mejia (2015) afirman: “Hoy en día cualquier computadora conectada a internet está expuesta a diversas amenazas. Una consecuencia es el aumento en el número de ataques informáticos”. Cuando un usuario accede a un sitio web, éste tiene cierto tipo de acceso a los diferentes recursos que se encuentran conectados a éste para su correcto funcionamiento como: los mismos servidores web, servidores de BD, etc.; tomando las medidas de control necesaria se puede proporcionar un entorno donde los usuarios puedan trabajar de forma segura y sin preocupaciones de algún tipo ataque.

Los ataques a sitios web son unos de los objetivos más atractivo para los atacantes, algunas razones son las siguientes:

- **Disponibilidad y accesibilidad. -** Los sitios o aplicaciones web se encentran disponibles en cualquier momento del día, tanto en el día como en la noche; además de que estos se encuentran de forma pública, por lo que cualquier usuario tiene acceso a éste.
- **Familiaridad.** - Actualmente la mayoría de personas tienen conocimientos sobre alguna interfaz web, obtener un navegador web es tan fácil además de que es un programa que se lo utiliza mucho; por lo que existen una gran variedad de herramientas de hacking que permiten acceder y comprometer este tipo de aplicaciones.
- **Facilidad.** - Para obtener una buena seguridad la configuración de un sitio o aplicación web para el uso público es sumamente compleja, por ende, no siempre se llega a tener una buena seguridad y los atacantes aprovechan algunas de estas deficiencias y logar penetrar el sitio alojado.
- **Publicidad.** - La publicidad es uno de los métodos llamativos para los atacantes, el solo hecho de modificar un sitio web tiene un valor para aquellas personas que se dedican a éste tipo de labores, ya que tienen como objetivo lograr algo que pocas personas pueden hacer.

1.2.2.3. Elementos de la seguridad

Los tres tipos de elementos que se deben proteger de un posible ataque informático son:

- **Software**

  Se entiende o se hace referencia a todo el conjunto de programas lógicos, parte intangible de un ordenador, como puede ser: Sistema Operativo (SO), hojas de cálculos, etc.

- **Hardware**

  Son los componentes físicos que conforman un ordenador, parte tangible, como puede ser: CPU, cables, medios de almacenamientos, etc.

- **Datos**

  Consiste en la información lógica que es manejado por el software y hardware como: datos que se envíen por medio de la red, almacenamiento de una Base de Datos (BD).

Por lo general, los datos es el elemento principal que se debe pensar en proteger, ya que sin las debidas precauciones éstos se podrían perder y sería difícil de recuperar; además de que es el objetivo principal en toda amenaza.

1.2.2.4. Amenaza

Las amenazas se dividen en 4 tipos:

- **Interrupción**

  Consiste en hacer que un objeto del sistema, al cual se atacará, se pierda o quede no disponible; ocasionando una pérdida o reducción de la disponibilidad en el sitio.

- **Intercepción**

  Es cuando un elemento no autorizado logra acceder de cierta forma a un determinado objeto del sistema que se encuentra en red, comprometiendo la

confidencialidad de la organización y a su vez, permitiendo que; éste se modifique.

- **Modificación**

Afecta a la integridad, el cual, además de poder acceder a un objeto o entidad de forma no autorizada, se consigue modificarlo y talvez dejarlo inutilizable.

- **Fabricación**

Se origina cuando se introduce un objeto o entidad falsificados al sistema

1.2.2.5. Mecanismos

IPA (2011) manifiesta: “Para mantener la seguridad de su sitio web, se necesita tomar las medidas de seguridad apropiadas en cada componente del sitio web”. Al momento de tomar alguna medida de seguridad, hay que tener en cuenta los siguientes mecanismos:

- **Prevención**

Consiste en todos aquellos mecanismos que permiten mejorar la seguridad de un determinado sistema o aplicación mientras se encuentra funcionando de forma normal, permitiendo prevenir posibles violaciones al sistema; un ejemplo claro sobre éste mecanismo es cuando se hace uso del cifrado por medio de la transmisión de información, ya que de ésta forma se evita que un posible atacante llegue a escuchar las conexiones.

- **Detección**

Éste trata de todos aquellos mecanismos que permiten detectar o monitorear las posibles violaciones o intentos de violación que pueden ocurrir en un determinado sistema, un ejemplo de éste mecanismo son los sistemas de auditorías.

- **Recuperación**

Consiste en aquel mecanismo que se aplica en el momento que se da o se detecta una violación, el cual permite retornar a un estado anterior y a un correcto

funcionamiento, un ejemplo claro es cuando se realizar el backup de un sistema para prevenir futuros inconvenientes. Además de lo ya dicho, éste mecanismo posee como tras fondo, ya que no solo basta con retornar a un estado anterior y dejar todo funcionando, sino que analizar cuál fue el alcance y el agujero por donde se realizó la violación; de ésta forma se puede prevenir futuros ataques mejorando la seguridad.

A pesar de que los 3 mecanismo, descaritos anteriormente, son importantes para mantener una buena seguridad en el sistema, se hace más énfasis en los mecanismos de prevención y detección; por lo que detectar o evitar una posible violación es mucha más productivo y eficaz que dejar que primero suceda y después tratar de controlarlo, ya que hasta q se encuentre una solución al problema, la información que maneja dicha entidad se podría ver comprometida.

Dentro de los mecanismos de prevención más comunes se tiene:

- Mecanismos de autentificación e identificación
- Mecanismos de control de acceso
- Mecanismos de separación
- Mecanismos de seguridad en las comunicaciones

1.2.2.6. Ataques Remotos

Existen una gran infinidad de ataques, dentro de estos se tiene:

- **Escaneo de puertos**

  Por lo general esta es una de las actividades que el atacante realiza a su objetivo, obteniendo información básica como el sistema operativo instalo en la máquina, arquitectura de la red, etc. Si el atacante descubre que ciertos puertos se encuentran abiertos, éste podría ingresar al equipo.

  Una de las aplicaciones más utilizadas para éste tipo de proceso es nmap, permite realizar el proceso de forma cómoda y brinda información necesaria.

- **Spoofing**

Consiste en la creación de tramas TCP/IP, donde el atacante simula tener la identidad de un equipo que se encuentre en la red donde hace uso de una dirección IP falsa, consiguiendo tener acceso a ciertos recursos que se encuentra compartido con el ordenador suplantado. Éste ataque se encuentra constituido de 3 componentes, el atacante, el atacado y el equipo suplantado; dónde el atacante para poder llevar a cabo su plan realiza una comunicación falsa con el equipo central que le comparte los recursos.

- **Negaciones de servicio**

Su propósito consta en denegar el acceso a un determinado equipo, en el cual los agresores evitan que los diferentes usuarios puedan tener acceso a éste, ocasionando que el servidor no se encontrará disponible por un tiempo determinado.

- **Interceptación**

Trata en capturar tramas que se encuentran circulando en la red mediante una aplicación que se encuentra instalada en un determinado equipo.

1.2.2.7. Ataques web

Mitchell (2011) afirma: “Para sistemas como servidores que están diseñados para estar "siempre encendidos", la seguridad es un problema importante”. Los servidores web al estar siempre disponible, por lo que nunca se apagan, se han convertido en el blanco principal para los atacantes; dentro de los ataques más comunes que ocurren a un servidor web se tiene:

- **SQL Injection**

Es uno de los ataques, hasta ahora, más usados, el cual por medio de la modificación de una cadena de consulta se puede explotar una posible vulnerabilidad, permitiendo al atacante pode ingresar a la información que se encuentra en el BD.

Al tener acceso al BD el atacante podría:

- ✓ Ver información confidencial
- ✓ Falsificar o eliminar registros
- ✓ Omitir la autentificación de inicio de sesión, etc.

- **Ataque de inyección de comandos de SO**

Al brindar una ventana de comando en ciertos aplicativos webs, le dan oportunidad al atacante de poder ejecutar comandos a nivel de Sistema Operativos, dejando al sistema vulnerable.

Al momento de ejecutar este ataque, el atacante podría

- ✓ Ver, falsificar y eliminar archivos del servidor, manipular de forma maliciosa el sistema
- ✓ Descargar e instalar programas maliciosos
- ✓ Convertir al sistema en un punto de partida para infectar y acceder a otros servidores o equipos.

- **Ataque de directorio transversal**

Ciertas aplicaciones web otorgan un acceso casi completo al servidor web, permitiendo especificar los nombres de os diferentes archivos que se encuentran en el servidor, si la aplicación no se encuentra con una debida seguridad, el atacante podría colocar un archivo malicioso, ocasionando que el sistema realice operaciones involuntarias.

La consecuencia de dicho ataque podría ser:

- ✓ Difundir información sensible
- ✓ Eliminar o modificar los diferentes archivos de configuración

- **Secuestro de sesión**

La mayoría de aplicaciones contiene un sistema de logueo para que el usuario pueda acceder a administrarla, al momento de iniciar sesión se crea un ID el cual permite identificar al usuario y administrar las diferentes sesiones. Si en el proceso no se crea debidamente el ID, el atacante podría robarla de un usuario y hacerse pasar por él, obteniendo acceso a lo que sería cierta información que maneja el sistema.

Si el atacante tiene éxito con la explotación de la sesión errónea, podría realizar acciones a las cuales tiene acceso dicho usuario:

- ✓ Acceder a los diferentes servicios que se brinda y modificarlos con el fin de obtener beneficios económicos
- ✓ Modificación de información
- ✓ Acceder a información personal de la organización

- **Ataque de scripting entre sitios**

Algunas aplicaciones web generan el contenido de su página en relación a una entrada del usuario, un ejemplo claro de esto puede ser los resultados de búsquedas, anuncios, etc.; si éste proceso no es regulado con alguna función de seguridad el atacante podría colocar scripts arbitrarios con el fin de realizar acciones maliciosas.

Si se aprovecha ésta vulnerabilidad el atacante podría:

- ✓ Obtener información confidencial por medio de una página falsa
- ✓ Si los datos ingresados por el usuario se almacenan en cookies, éste los puede obtener desde el navegador y revelar su información.

- **Ataque de falsificación de solicitudes entre sitios**

Para poder realizar alguna transacción o proceso e algunos sitios web, es necesario iniciar sesión, y si el sistema no posee algún mecanismo que verifique si la solicitud que se realiza fue hecha por el usuario o por algún agente externo, el sitio web podría aceptar cualquier tipo de solicitud que sea configurada por atacantes externos perjudicando al usuario.

Si éste ataque llegara a pasar, el atacante podría:

- ✓ Acceder a los servicios que se encuentran disponibles para el usuario (transferencias, compras, etc.).
- ✓ Modificar información confidencial a la cual solo el usuario tendría acceso.

- **Ataque de inyección de encabezado de correo**

Algunas aplicaciones web proporcionan un servicio de envíos de correo electrónicos a las diferentes direcciones de sus usuarios, éstas direcciones se encuentran solamente bajo la supervisión de un administrador. Si la aplicación no posee una correcta gestión de seguridad el atacante podría acceder a éstas y modificarlas.

Si el atacante llega a tener acceso a ésta información, podría comenzar a enviar spam con fines maliciosos a los diferentes correos electrónicos de los usuarios.

1.2.3. Logs

Los archivos logs son registros en el cual se almacenan los diferentes eventos que se realizan dentro de un sistema, red o aplicación. Se encuentran compuestos de entradas, dónde cada una de éstas representan alguna información sobre algún evento o suceso que haya ocurrido en el sistema; anteriormente éste tipo de archivo se los usaba con el fin de solucionar problemas, pero con el paso del tiempo se ha visto que posee una infinidad de uso como optimizar el rendimiento de un sistema, registrar las diferentes acciones que realice un usuario, proporcionar información para detectar actividades maliciosas, etc.

Langhnoja, Barot, y Mehta (2012) manifiestan: “Los archivos logs de la web son archivos que contienen información sobre la actividad de los visitantes del sitio web. Los archivos de registro son creados por los servidores web automáticamente. Cada vez que un visitante solicita cualquier archivo (página, imagen, etc.) de la información del sitio en su petición se adjunta a un archivo de registro actual”.

Éstos tipos de archivos posee un gran impacto en las organizaciones o empresas, ya que actualmente la tecnología ha innovado muchas de las labores que comúnmente las personas realizan, además las empresas hacen uso de equipos informáticos para llevar a cabo sus labores como son: almacenar su información en las Base de Datos, ofrecer productos y servicios por medio de la web, etc.

Debido a esto el número de amenazas contra éstos dispositivos ha aumentado considerablemente, creando una gran necesidad en la administración de registro en la seguridad informática, el cual consiste en generar, almacenar, analizar y eliminar los diferentes datos en los registros de un computador.

1.2.3.1. Fundamentos de los registros de seguridad informática

> Los mensajes de los logs son muy importantes para los administradores de sistemas para entender el comportamiento previo del sistema. Pueden mostrar pistas de lo que está pasando en el sistema, en las redes, así como recolectar datos sobre el rendimiento y sobre todo la detección de intrusiones. Por lo tanto, son una gran fuente de información que permite determinar lo que ha pasado después de un incidente para poder tomar las medidas oportunas. (Mellouk, 2016)

Sin importar las circunstancias, los archivos logs poseen una gran relevancia en la administración de una organización en la seguridad informática, un ejemplo de éstos puede ser los logs de los dispositivos de red o de algún software de monitoreo, la información que éstos equipos registran pueden ser muy útiles a la hora de realizar alguna operación o alguna auditoría, permitiendo demostrar si se cumplen o no las regulaciones de la empresa.

Sin embargo, cuando de equipos se trata, la información que éstos almacenan se la utiliza como información complementaria.

Los tipos de registros a considerar son los siguientes:

- **Seguridad de software**

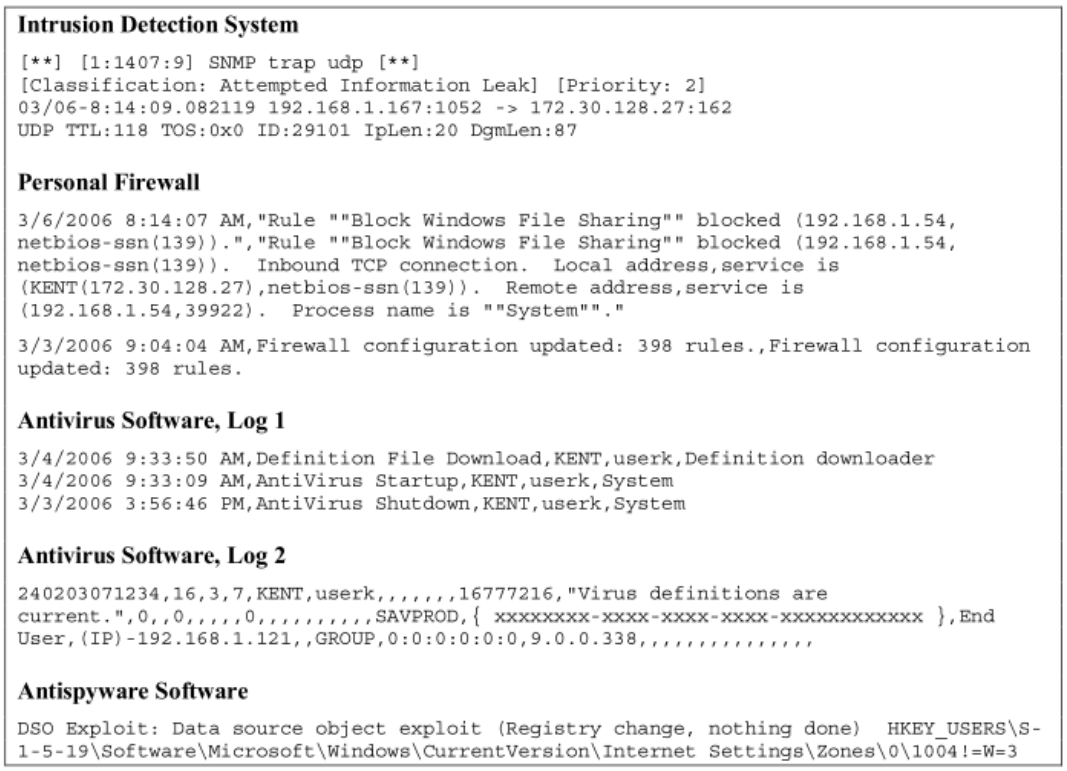
  Normalmente las organizaciones hacen uso de algún software de seguridad para proteger la información almacenada y controlar alguna actividad maliciosa, Éste tipo de sistema se encuentran conformado por:

  - ✓ **Software antimalware. -** Una representación de éste tipo de software son los antivirus, ya que éstos registran todo aquel suceso que se encuentre relacionado con un malware, además crea archivos logs de las exploraciones que éste realice.
  - ✓ **Sistema de detección y prevención de intrusos.** - Los archivos logs de éste tipo de sistema registran información de forma detallada sobre algún comportamiento sospechosos y de los ataques que se realicen;

algunos de estos sistemas se ejecutan de forma periódica, almacenando información en forma de lotes.

- ✓ **Software de acceso remoto.** - Ésta forma de seguridad se la realiza mediante el uso de redes virtuales o VPN, dónde estás registran los inicios de sesión que hayan tenido éxito y los que no, acompañada de cierta información complementaria como es la fecha y hora en la que el usuario se haya conectado o desconectado, la cantidad de datos que se enviaron y las que se recibe, etc.; algunos tipos de VPN almacenan información detallada sobre los recursos que se hayan utilizado.
- ✓ **Web proxies.** - Consisten en servidores intermedios por los cuales se realiza el acceso algún sitio web, realizando solicitudes de las páginas accedidas en nombre de los usuarios y registrando copias de los sitios web recuperados en la caché, con el fin de que el acceso sea más eficiente. Pero otro uso de éste tipo de sistema es la de restringir el acceso web o agregar, entre el cliente y el servidor, una capa de protección.
- ✓ **Software de gestión de vulnerabilidades.** - Se encuentra constituido por un software de administración de parches y de evaluación de vulnerabilidades, se encarga de registrar información acerca de los parches de instalación y de la vulnerabilidad de algún equipo determinado.
- ✓ **Servidores de autentificación.** - Registran información detallada de cuando se realice alguna autentificación, ésta información se encuentra constituida por el nombre del usuario, fecha, hora, si hubo falla o no y su nombre de origen.
- ✓ **Enrutadores.** - Permiten tener un control del tráfico de red, almacenando información sobre el tráfico y las actividades que se realizan.
- ✓ **Cortafuegos.** - Al igual que un enrutador, permiten tener un control del tráfico realizado, pero de una forma más completa y detallada; permitiendo rastrear el origen del tráfico y el contenido a cuál se accedió.
- ✓ **Sistema de cuarentena de red.** - Permiten realizar una comprobación y llevar un registro de cómo se encuentra la seguridad de un

determinado servidor antes de publicarla en la red, dónde si éste no aprueba los diferentes controles que se realizan, pasan a un estado de cuarentena.



Intrusion Detection System

```
[**] [1:1407:9] SNMP trap udp [**]
[Classification: Attempted Information Leak] [Priority: 2]
03/06-8:14:09.082119 192.168.1.167:1052 -> 172.30.128.27:162
UDP TTL:118 TOS:0x0 ID:29101 IpLen:20 DgmLen:87
```

Personal Firewall

```
3/6/2006 8:14:07 AM,"Rule ""Block Windows File Sharing"" blocked (192.168.1.54,
netbios-ssn(139)).","Rule ""Block Windows File Sharing"" blocked (192.168.1.54,
netbios-ssn(139)).  Inbound TCP connection.  Local address,service is
(KENT(172.30.128.27),netbios-ssn(139)).  Remote address,service is
(192.168.1.54,39922).  Process name is ""System""."

3/3/2006 9:04:04 AM,Firewall configuration updated: 398 rules.,Firewall configuration
updated: 398 rules.
```

Antivirus Software, Log 1

```
3/4/2006 9:33:50 AM,Definition File Download,KENT,userk,Definition downloader
3/4/2006 9:33:09 AM,AntiVirus Startup,KENT,userk,System
3/3/2006 3:56:46 PM,AntiVirus Shutdown,KENT,userk,System
```

Antivirus Software, Log 2

```
240203071234,16,3,7,KENT,userk,,,,,,,16777216,"Virus definitions are
current.",0,,0,,,,,0,,,,,,,,,,,SAVPROD,{ xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx },End
User,(IP)-192.168.1.121,,GROUP,0:0:0:0:0:0:0,9.0.0.338,,,,,,,,,,,,,,
```

Antispyware Software

```
DSO Exploit: Data source object exploit (Registry change, nothing done)  HKEY_USERS\S-
1-5-19\Software\Microsoft\Windows\CurrentVersion\Internet Settings\Zones\0\1004!=W=3
```

Ilustración 6 Ejemplos de registro logs de software de seguridad

Fuente: (Murugiah, 2006)

- **Sistemas operativos**

Los Sistemas Operativo almacenan una gran cantidad información relacionada con la seguridad, dentro de los tipos de datos que éstos guardan se tiene:

- ✓ **Eventos del sistema. -** A ciertas acciones operacionales como apagar el sistema o iniciar algún servicio se les denomina eventos de sistema; existen alguno Sistemas Operativos que permiten especificar el tipo de evento que se puede registrar, pero generalmente se almacenan aquellos que fallen o los que se logren realizar con éxito.
- ✓ **Registros de auditoría.** - Éste tipo de registro contiene información basado en la seguridad como la autentificación de un usuario, cuando se accede algún archivo, modificaciones a la política de seguridad, etc.

Metaute (2013) afirma: “Lo que se consigue con una auditoría de seguridad es garantizar a los usuarios que sus datos y sistemas están siendo protegidos con las medidas de seguridad apropiadas generando confianza en la compañía, entidad o propietario de la aplicación”. Éste tipo de registro posee un gran punto a favor, ya que permite identificar alguna posible actividad maliciosa que se puede dar en un equipo, dónde una vez confirmado la actividad sospechosa se realizan consultas al sistema para obtener más información de la actividad; por ejemplo, si se detecta alguna actividad sospechosa hacia un determinado equipo desde algún dispositivo de seguridad de red, se podría conocer si un usuario ha iniciado sesión en el momento del ataque y si éste tuvo éxito, mediante los registros almacenado en el sistema.

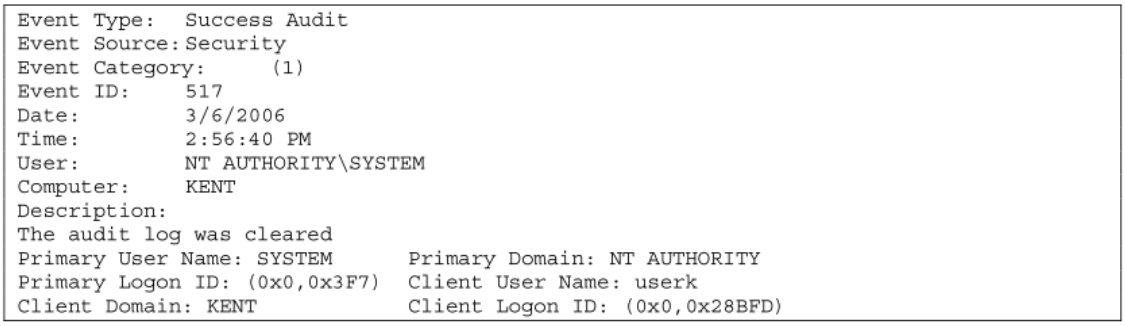


Ilustración 7 Ejemplo de registro del SO

Fuente: (Murugiah, 2006)

- **Aplicaciones**

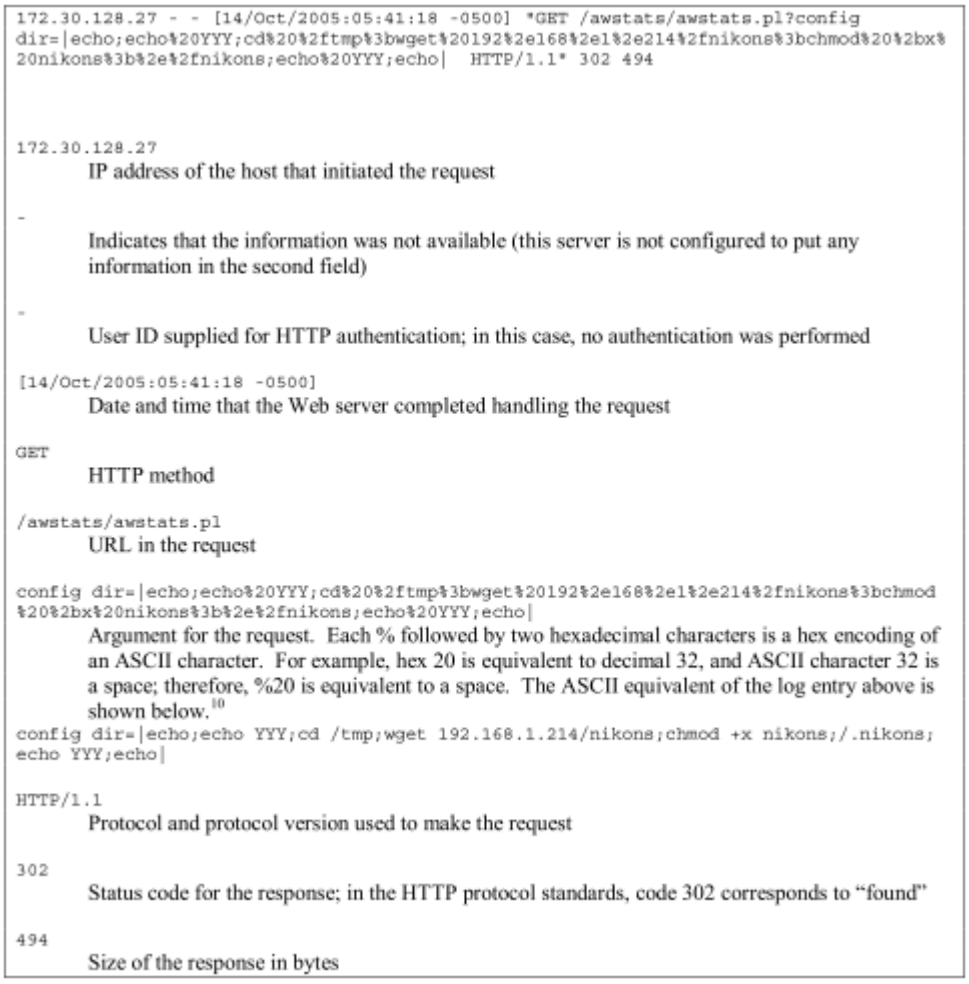
En la actualidad existen una gran variedad de formas de como violar o robar información de algún equipo, y para que esto no suceda, las organizaciones y personas en general hacen uso de aplicaciones basados en la seguridad para poder controlar éstos tipos de amenazas, ya que éstos hacen uso de otros recursos informáticos para poder realizar sus labores diarias, los cuales manejan información confidencial.

> La carrera por ser el primero en adaptar estas tecnologías, junto con la falta de experiencia, dejadez o malas prácticas por parte de los desarrolladores, hace que en ocasiones se incurran en errores que pueden dar lugar a vulnerabilidades de seguridad de distinta gravedad. (Metaute, 2013)

Algunas de éstas aplicaciones fueron desarrolladas para poder manejar sus propios archivos logs, mientras que otras hacen uso de los archivos logs del sistema; los registros de éstas aplicaciones varían, ya que cada una posee una estructura de cómo se almacenan los datos, dentro de los tipos de registros más comunes se tiene:

- ✓ **Solicitudes de clientes y respuesta del servidor. -** Consiste en poder identificar incidentes y monitorear el uso del sistema, dónde si la autenticación del usuario fue exitosa, se pude conocer la identidad del usuario, el tipo de respuesta enviada por el servidor, aplicaciones que se accedió, etc.
- ✓ **Información de la cuenta.** - Permite conocer si la autenticación de usuario fue exitosa o fallida, modificaciones en la cuenta, privilegio que posee como usuario, etc.; en sí, identifica si un usuario ha usado el sistema y cuando lo hizo.
- ✓ **Uso de información. -** Permite monitorear la cantidad de procesos realizado en determinado momento y su tamaño.
- ✓ **Acciones operacionales significativas. -** Consiste en poder identificar cuando ocurre una falla en una determinada aplicación, cambios, inicios y cierres en la misma.

Existen algunas aplicaciones que no se valen de otras para poder crear los registros, por lo que los archivos logs de ésta se vuelven valiosos para temas de seguridad de la misma. En algunos casos éste tipo de registros son complicados de utilizar, ya que como se encuentra en el formato definido por el propietario, es muy poco probable poder encontrar un método flexible para analizarlos.



```
172.30.128.27 - - [14/Oct/2005:05:41:18 -0500] "GET /awstats/awstats.pl?config
dir=|echo;echo%20YYY;cd%20%2ftmp%3bwget%20192%2e168%2e1%2e214%2fnikons%3bchmod%20%2bx%
20nikons%3b%2e%2fnikons;echo%20YYY;echo|  HTTP/1.1" 302 494
```

172.30.128.27
IP address of the host that initiated the request

-
Indicates that the information was not available (this server is not configured to put any information in the second field)

-
User ID supplied for HTTP authentication; in this case, no authentication was performed

[14/Oct/2005:05:41:18 -0500]
Date and time that the Web server completed handling the request

GET
HTTP method

/awstats/awstats.pl
URL in the request

config dir=|echo;echo%20YYY;cd%20%2ftmp%3bwget%20192%2e168%2e1%2e214%2fnikons%3bchmod
%20%2bx%20nikons%3b%2e%2fnikons;echo%20YYY;echo|
Argument for the request. Each % followed by two hexadecimal characters is a hex encoding of an ASCII character. For example, hex 20 is equivalent to decimal 32, and ASCII character 32 is a space; therefore, %20 is equivalent to a space. The ASCII equivalent of the log entry above is shown below.[10]

```
config dir=|echo;echo YYY;cd /tmp;wget 192.168.1.214/nikons;chmod +x nikons;/.nikons;
echo YYY;echo|
```

HTTP/1.1
Protocol and protocol version used to make the request

302
Status code for the response; in the HTTP protocol standards, code 302 corresponds to "found"

494
Size of the response in bytes

Ilustración 8 Archivo log de un servidor web

Fuente: (Murugiah, 2006)

1.2.3.2. La necesidad de gestión de logs

Actualmente la mayoría de sistemas de información y comunicación dejan evidencias de su estado, operación y resultados en forma de logs con el objetivo de ofrecer a administradores, desarrolladores, operadores y usuarios información detallada respecto a su funcionamiento. Esta información procesada a través de los mecanismos adecuados podría convertirse en una base de datos de eventos con utilidad en diversos fines, entre los cuales se encuentran: Administración de recursos, detección de intrusiones, la resolución de problemas, análisis forense y auditorías. (Carrión, 2015)

La gestión de logs es un gran punto a favor para la mayoría de las empresas, ya que permite garantizar que los logs almacenen información con los detalles suficientes y necesarios para una correcta administración, esto ayuda a identificar los incidentes, actividades maliciosas, problemas operativos, etc.; y a su vez proporcionan información para resolver ciertos problemas; también bridan un gran aporte en cuanto auditoría, análisis forense y establecimiento de líneas base para el análisis de problemas a largo plazo.

Cabe recalcar que en la actualidad existen leyes que obligan a las instituciones a trabajar con éste tipo de archivos.

1.2.3.3. Desafíos en la gestión de logs

A pesar de que los archivos logs brinda múltiples beneficios para la administración de una organización, llevar a cabo la gestión presenta algunos problemas, éstas se la pueden dividir en 3 partes:

- Problemas al momento de generar un archivo log debido a la gran variedad que éstos se conforman y a la forma de aprovecharlos.
- Los 3 componentes fundamentales como confidencialidad, integridad y disponibilidad podrían verse comprometidos.
- El personal responsable de realizar su debido análisis, podría no tener la experiencia suficiente al momento de trabajar con éste tipo de archivos.

Para un mejor entendimiento se hablará de los siguientes temas:

- **Generación de registro y almacenamiento**

  La generación de los Sistemas Operativos, aplicaciones, equipos, entre otras cosas; generan y almacenan una infinidad de archivos log, complicando su administración de las siguientes formas:

  - ✓ **Gran cantidad de registros. -** Una institución se encuentra conformada por una gran variedad de equipos, de los cuales cada uno de éstos generan sus propios archivos de registros; la administración y análisis de éstos resultaría complicada debido a la abundancia de los datos.

- ✓ **Contenido de registro inconsistente.** - Cada equipo crea su sistema de archivos log, y el formato o estructura que éstos lleguen a tener se encuentra definido por el propietario o el desarrollador; muchas de éstas solo almacenan información relevante, dejando a un lado los demás datos, por lo que, si la información obtenida de un equipo no posee ninguna relación con la información de otro, no se podría analizar nada por la inconsistencia de éstos. Por ejemplo, en el archivo log1 se registra la IP de un equipo, pero no el usuario y en el archivo log 2 se registró el usuario, pero no la IP del equipo, no habría como relacionar si la IP del log1 pertenece al usuario del log2.
- ✓ **Marca de tiempo inconsistente.** - Cuando se genera un registro, éste se almacena con la hora que posee el equipo internamente; por lo que, si el tiempo del equipo es inexacto, éste se guarda con la hora inexacta complicando el análisis del registro.
- ✓ **Formato de registros inconsistentes. -** Como se ha hablado anteriormente, cada registro maneja su propio formato, unos utilizan el texto separados por coma, bases de datos, formato de diferentes (XML), archivos binarios, fácil entendimiento para las personas, formato estándar o formato propietario, algunos se almacenan de forma local y otros se transmiten a otro sistema para su procesamiento, etc.

Debido a todas éstas complicaciones, las organizaciones necesitan implementar métodos automatizados para poder convertir los diferentes registros a un único formato para un mejor entendimiento.

- **Protección de log**

Los archivos log contienen información confidencial como contraseñas de un usuario, información de correos electrónicos, etc.; por lo que deben estar protegidos para que personas con intenciones maliciosas no puedan acceder a estos. Además de la seguridad de estos archivos, también es necesarios asegurar la disponibilidad que éstos tengan, ya que por lo general estos archivos almacenan un máximo de 10.000 eventos o 100 megabytes de datos y sí se llega al límite,

éstos podrían detener el almacenamiento de registros o sobrescribir registros anteriores, por lo que se perdería la disponibilidad e información ya almacenada.

- **Análisis de log**

  > Para dar soporte a los procesos de gestión y monitoreo, es importante contar con herramientas que de manera automática permitan consolidar, filtrar, consultar y analizar la información registrada en los logs, como es el caso de los Sistemas de Gestión de eventos e Información de Seguridad (SIEM) de la misma manera se deben establecer responsables que definan políticas de retención, desecho y preservación de los mismos. (Mejía, 2014)

  Generalmente, en las organizaciones, quienes se encargan de realizar el análisis de registros logs son los administradores de sistemas y redes, en el cual por medio de las entradas del registro identifican los eventos de interés; ésta gestión de análisis ha sido tratada con un tema de poca prioridad, ya que otras actividades como dar solución a vulnerabilidades o solución a problemas operacionales requieren de una respuesta rápida, dejando de lado al análisis de éstos registros. Además, el personal encargado de ésta gestión no recibe una buena capacitación de cómo llevar a cabo el análisis, ni de las herramientas suficientes para automatizar y facilitar el proceso.

Muchas de éstas herramientas son muy útiles, por lo que ayudan a encontrar patrones, que para las personas se les hace difícil de encontrar, y relacionar diferentes registros almacenados pertenecientes a un mismo evento, brindando una información más detallada del suceso; cabe mencionar que muchos consideran al análisis de logs como una pérdida de tiempo, por la gran cantidad de tiempo y paciencia que se tiene q invertir para su debido análisis.

1.2.3.4. Como llevar a cabo la gestión de análisis

A pesar de que llevar acabo la gestión de análisis de archivos logs presenta algunas dificultades, existen algunos métodos para realizar resolver algunos de éstos obstáculos como:

- **Priorizar la gestión de registro**

Definir sus objetivos para el almacenamiento de registros necesarios, con el fin llevar un correcto monitoreo de éstos, incluyendo las políticas organizacionales.

- **Establecer políticas y procedimiento**

Brindan un gran beneficio a la organización, por lo que permiten garantizar que se cumplan las leyes estipuladas por la misma; un ejemplo de esto puede ser las auditorias, ya que permiten verificar que se cumpla las leyes y pautas de los registros.

- **Mantener una infraestructura segura**

Cuando una organización posee una buena infraestructura de registro, es mucho más fácil preservar la integridad de los mismos ante cualquier ataque o acceso no autorizado.

- **Brindar soporte para el personal encargado de los logs**

Las organizaciones deben proporcionar la capacitación y formación necesaria para el personal responsable de la gestión de logs, además de las herramientas, para que éstos adquieran conocimientos y habilidades necesarios de acuerdo a sus labores operacionales.

1.2.3.5. Arquitectura

La arquitectura de archivos logs está conformada por:

- **Generación de logs**

Éste primer nivel se encuentra conformado por los diferentes equipos que generan los registros, algunos de éstos ejecutan aplicaciones o algún servicio de logueo de clientes, dónde los datos almacenados se encuentran disponibles por medio de la red.

- **Análisis y almacenamiento de logs**

El segundo nivel lo conforman aquellos servidores que reciben los archivos los provenientes de los equipos de primer nivel, éstos archivos se transiten en un horario establecido o cuando existe cierta cantidad de datos.

- **Monitoreo de logs**

Por último, el tercer nivel se encuentra con aquellas consolas que permiten visualizar los datos que se encuentran en los archivos y generar los respectivos informes, en algunas infraestructuras las consolas de monitoreo se usan para gestionar los registros de los servidores de almacenamiento y de los equipos clientes.

CAPÍTULO II

2. Metodología

2.1. Descripción de la investigación a realizar

Se procedió a la implementación de una herramienta de análisis de archivos logs, donde ésta comenzó a analizar todas las acciones y eventos que se realizaban en un servidor web.

2.2. Tipo de investigación

Los tipos de investigación en los que se basó éste trabajo son:

- **Según el objetivo de la investigación**

Investigación aplicada. - Consiste en una investigación que se centra en la búsqueda de mecanismos para logar o alcanzar un objetivo en común, en este caso se realizó una búsqueda de programas que permita leer de mejor forma los archivos logs y resaltar su la utilidad de los mismos.

- **Según el nivel de profundización en el objeto de estudio**

Investigación exploratoria. - Trata de una investigación a profundidad de aspectos que, en cierta forma, aún no han sido completamente tan tratados; en este caso se investigó sobres los archivos logs, su estructura y la información que estos almacenan.

Investigación descriptiva y explicativa. - Descriptiva porque se detalló los diferentes fenómenos que pueden ocurrir al momento de que el servidor reciba un determinado ataque o se realice un evento; y explicativa porque se analizaron los archivos logs obtenido y se explicaron las causas y consecuencias de las acciones realizadas para la obtención de los resultados.

- **Según el grado de manipulación de variables**

Investigación Cuasi-experimental. - Se obtuvo el control y manipulación de ciertas variables y, dependiendo de estas se obtuvieron diferentes resultados para proceder a su análisis, un ejemplo puede ser los archivos logs del servidor, se hizo uso de estos, pero no se tuvo una manipulación completa de los registros que estos guardan, ya que estos de por si tienen establecido su formato.

- **Según el tipo de inferencia**

Investigación inductiva. - Permite concluir mediante la obtención de hechos que tendrán como base la realización práctica de la tesis, como es la implementación de herramientas de análisis de log de un servidor web y concluir con la relevancia que tiene la información que estos archivos almacenan.

- **Según el tipo de dato empleado**

Investigación cuantitativa. - Permite concluir mediante la obtención de gráficos estadístidos generados por la herramienta analizadora de logs.

2.3. Métodos y técnicas

El presente proyecto se enfocó en los siguientes métodos y técnicas:

Técnica de escalas de medición

> Todo problema de investigación científica, aún el más abstracto, implica de algún modo una tarea de medición de los conceptos que intervienen en el mismo, porque si tratamos con objetos como una especie vegetal o un comportamiento humano nos veremos obligados ya sea a describir sus características o a relacionarse éstas con otras con las que pueden estar conectadas. (Behar, 2010)

Técnica de escalas de medición ofreció un gran aporte en la realización del proyecto, ya que permitió describir un comportamiento del problema u objeto investigado y las posibles relaciones que puede haber con otro objeto similar.

Técnica de observación

Behar (2010) manifiesta: “La observación consiste en el registro sistemático, válido y confiable del comportamiento o conducta manifiesta. Puede utilizarse como instrumento de medición en muy diversas circunstancias”. Esta técnica determinó como es el comportamiento del servidor web al momento de realizar los diferentes ataques y eventos, para luego proceder al procesamiento de la información que este guarda.

Método lógico deductivo

Este método permitió dar a conocer las diferentes consecuencias que se desconocen mediante la lectura de los archivos logs. Según Behar (2010) dice: “Sirve para descubrir consecuencias desconocidas, de principios conocidos”. Por lo que si en los registros guardados se llega apreciar cierto direccionamiento que aparece de forma muy seguida, se podría deducir que podría tratarse de un ataque.

Método inductivo

Abreu (2014) manifiesta: “Mediante este método se observa, estudia y conoce las características genéricas o comunes que se reflejan en un conjunto de realidades para elaborar una propuesta o ley científica de índole general”. Esta metodología se lleva de la mano con la técnica de observación, ya que ésta consiste en la obtención de conclusiones a partir de un acontecimiento dado y por medio de la observación se será testigo del comportamiento del servidor web al momento de realizar el ataque.

2.4. Descripción del instrumento

Los instrumentos que permitieron llevar a cabo la investigación son:

La observación

Permitió llevar una planificación de los procesos que se plantearon, ésta se compuso de:

- Tareas que se realizaron, ver **anexo1**
- Plazo o cronograma que se realizó, ver **anexo 1**
- Lista guía de los aspectos que se investigaron, ver **anexo2**

Estableciendo de por sí, los objetivos y aspectos que se trataron en la investigación; además de que ésta herramienta ayudo en el logro de obtener conocimientos o conclusiones ajustadas a la realidad.

Escalas de medición

Permitió describir el comportamiento del fenómeno se investigó, como fue el caso de un servidor web, del cual se logró entender de mejor forma el por qué de las cosas en el momento que se realizó una determinada acción.

2.5. Descripción de las técnicas de procesamiento y análisis

Los procesos que se efectuaron en la siguiente investigación fueron:

- **Técnicas de procesamiento**

Se realizó un plan o un cronograma que permitió llevar un orden de las distintas acciones que realizaron y materiales necesarios.

Se recopiló información de diferentes fuentes confiables, acerca de los ataques más comunes que se producen en este tipo de situación y de las herramientas para su respectivo análisis.

Se implementó una herramienta que facilite el análisis de los diferentes archivos logs generados en el servidor.

- **Técnicas de análisis**

Los datos se analizaron de forma gráfica, por lo que se visualizó en un gráfico estadísticos los diferentes registros de los eventos que se hayan almacenado en el archivo log, para un mejor entendimiento y análisis de los mismos.

2.6. Normas éticas

Para el presente proyecto se tomó en cuenta lo siguiente:

- No realizar plagio de otros archivos.
- No realizar actos indebidos de los datos que se procesen.
- Preservar la confidencialidad de la información obtenida.
- Realizar con total responsabilidad los procesos llevados.
- No manipular la información con fines para beneficios propios.

CAPÍTULO III

3. Resultados

3.1. Comparación y selección de la mejor herramienta de análisis

Para la implementación de la herramienta adecuada para obtener un análisis de logs más eficiente se realizó la comparación de dos herramientas llamas Graylog y ELK (Eleasticsearch-Logstash-Kibana), dónde se tomó como referencia las siguientes investigaciones para poder seleccionar la mejor de ellas:

<table>
<tr><th colspan="3">Primera comparación de herramientas de análisis de logs</th></tr>
<tr><td colspan="3">Monte (2016), en su investigación identifica cada uno de los componentes y características que se encuentran conformadas éstos sistemas como es la Base de Datos, como se recolecta los datos, interfaz gráfica de visualizaciones, Sistemas Operativos, etc.</td></tr>
<tr><th colspan="3">Base de Datos</th></tr>
<tr><td colspan="3">Debido a que las aplicaciones antes mencionadas poseen la misma base de datos, llamada ElasticSearch, no existe una comparación como tal; pero se puede decir que:
Posee un motor de búsqueda distribuido y con una capacidad multitendencia.
Permite crear única instancia del software a la Base de Datos y trabajar con múltiples usuarios.
Trabaja de forma concurrente.
Brinda un servicio más rápido en el momento de ejecutar una búsqueda.
Se encuentra conformado por un Clúster, el cual permite ejecutar varias instancias de forma paralela.</td></tr>
<tr><th colspan="3">Recolección de datos</th></tr>
<tr><td>Rsyslog</td><td>Logstash</td><td>Beats</td></tr>
</table>

• Es uno de los estándares más usados en el reenvió de alertas y registros. • Es una utilidad de software de código abierto utilizada en sistemas informáticos UNIX y similares a Unix. • Permite crear filtros basado en el contenido generado del equipo. • Opciones de configuración flexibles. • Permite él envió de varios registros perteneciente a un único fichero. • Para el envío y recepción, hace uso de ciertos estándares como RFC3164. • A causa de la utilización de éstas estándar, opciones de envíos se reducen, al igual de la utilización de la herramienta.	• Motor de recolección de datos de código abierto desarrollado por Elastic. • Permite unificar dinámicamente datos de fuentes dispares y normalizar los datos en destinos de su elección. • Permite transformar un registro a una gran variedad de entradas o filtros. • Permite modificar los filtros de una forma que se adapten a la necesidad de una empresa. • Requiere de una Máquina Virtual Java para su funcionamiento. • Sistema multiplataforma, que puede ser ejecutado en cualquier Sistema operativo. • Herramienta creada por la misma persona que desarrolló la Base de Datos. • Garantiza que no habrá pizca de incompatibilidad en la implementación del sistema.	• Serie de agentes ligeros de reenvío, de código abierto. • Desarrollados también por Elastic como una alternativa más ligera a Logstash. • Se divide en 4 tipos de archivos Beats llamados Packetbeat, Filebeat, Metricbeat y Winlogbeat. • Cada tipo de Beats posee una función determinada. • A pesar de ser una alternativa a Logstash, no brinda la misma potencia.
Visualización (Interfaz gráfica)		
Graylog	Kibana	

- Es una solución para visualizar registros almacenados en una base de datos Elasticsearch.
- Permite realizar consultas avanzadas sobre los datos almacenados, crear tablones con los resultados de las mismas o incluso generar alarmas ante determinados datos o ausencia de ellos.
- Para acceder a la interfaz se hace uso de un navegador.
- El usuario debe autentificarse primero antes de poder realizar alguna acción en el sistema.
- Se encuentra basada a sistemas pertenecientes a la plataforma Debian, Red Hat, etc.;
- hace uso de mongo, como Base de Datos adicional para almacenar las diferentes configuraciones que el usuario realice.

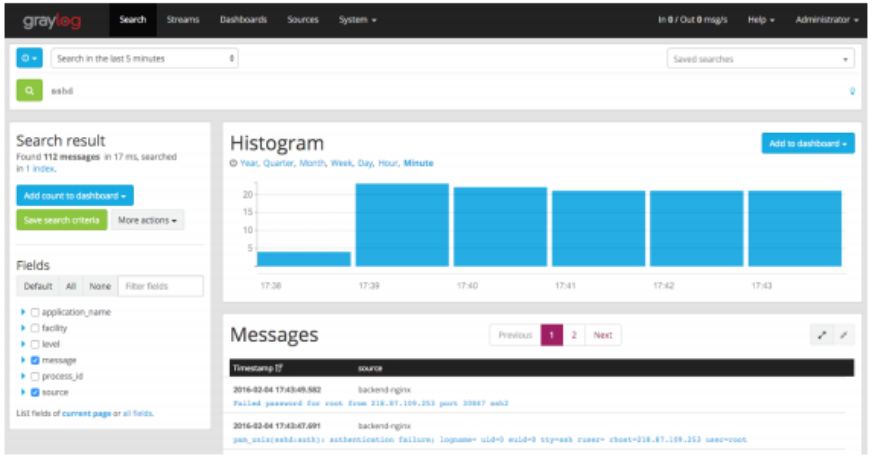


Ilustración 9 Interfaz de Graylog

- Potente herramienta pensada para el análisis masivo de datos almacenados en una base de datos Elasticsearch.
- Enfocada al Big Data.
- Ofrece una infinidad de nuevas posibles representaciones gráficas para los eventos, como mapas de calor, localizaciones, etc.
- Alternativa de Graylog, al igual que éste, se accede por medio de un navegador.
- No posee la opción de autentificación, pero se puede implementar por medio de un módulo.
- Sistema compatible con todos los sistemas operativos.

Ilustración 10 Interfaz de ELK

Selección
Una vez analizado cada uno de los componentes mencionados anteriormente, el autor Monte (2016) se inclina más hacia ELK, a causa de que es más personalizable que graylog; además de que es un sistema mucho más potente y a la vez compatible con múltiples plataformas. Una característica importante que hace resaltar más éste sistema a los demás, es que los otros sistemas se deben realizar filtros en el servidor que se quiere analizar para poder enviar los archivos logs, mientras que éste hace uso de un sistema de cola, permitiendo transformar los archivos recibidos en el servidor de análisis

Tabla 1 Comparación1 de herramientas de análisis

Fuente: (Monte, 2016)

Segunda comparación de herramientas de análisis de logs		
González (2015), en su investigación toma en cuenta 3 sistemas para el envío, recepción y análisis de archivos logs, estos son Syslog, Graylog y ELK		
Syslog	**Graylog**	**ELK**
• Es una familia de productos cuya principal función es enviar, recibir y almacenar entradas de logs. • Su arquitectura se basa en el modelo cliente-servidor. • El responsable de almacenar las diferentes entradas de archivos logs es el servidor, y el cliente se encarga de las operaciones de análisis. • Las operaciones que realiza suelen ocupar una gran cantidad de recursos, por lo que los clientes syslog pueden verse afectado en el rendimiento. • A pesar de brindar una gran variedad se técnicas para un envió, recepción y análisis eficaz; no destacan entre los sistemas similares.	• Se trata de uno de los productos de gestión de logs más completos existentes en el mercado. • Ofrece funcionalidades de monitorización, además del análisis y almacenamiento de mensajes. • Los protocolos por defectos que admiten son UDP, TCP y GELF. • No posee ninguna configuración por defecto para realizar los diferentes análisis, sino que el usuario debe añadir y personalizar el sistema para mostrar lo datos que se quieren analizar. • Se encuentra conformada por dos Base de Datos, Elasticsearch y mongoDB. • Elastic permite realizar las diferentes indexaciones y MongoDB almacena las	• ELK son las siglas de Elasticsearch-Logstah-Kibana y se denomina así a la arquitectura creada a partir de estos tres productos. • Al igual que graylog, hace uso de Elasticsearch como Base de Datos y de Kibana para poder visualizarlos. • Consiste en 3 herramientas diferentes que uniéndolas crean un sistema de gestión de logs potente y eficaz • A diferencia de Graylog permite un sinnúmero de protocolos para el almacenamiento de archivos logs, como UDP, TCP, GELF, RELP, etc; al igual varias fuentes como Base de Datos SQL y ficheros CSV

<table>
<tr>
<td>

- Para el correcto funcionamiento del sistema, se deben aplicar ciertos protocolos en ambos equipos (cliente y servidor).

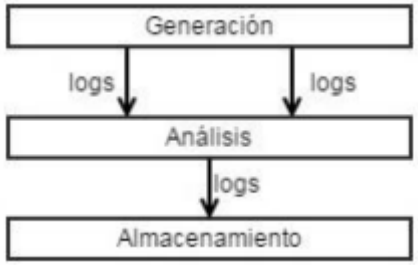


Ilustración 11 Arquitectura de Syslog

</td>
<td>

configuraciones de los usuarios y las estadísticas de los logs que se generan.

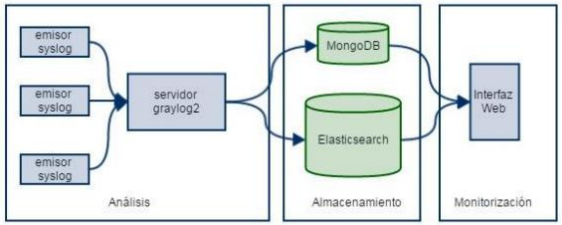


Ilustración 12 Arquitectura de Graylog

</td>
<td>

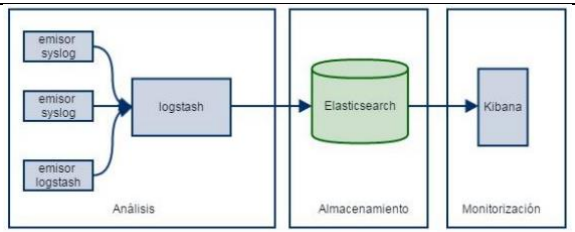


Ilustración 13 Arquitectura de ELK

</td>
</tr>
<tr>
<td colspan="3">Selección</td>
</tr>
<tr>
<td colspan="3">Finalizada la comparación del funcionamiento y arquitectura de las diferentes herramientas de análisis de logs, el autor González (2015) elije la herramienta ELK, ya que posee una herramienta de análisis, visualización y generación de reportes muy potente; además la Base de Datos Elasticsearch permite realizar búsquedas de forma eficiente en textos con o sin estructura, pudiendo manejar grandes volúmenes de datos. Su comunidad de soporte y desarrollo se encuentra muy activa, también de que se encuentra en múltiples lenguajes y soportada por todos los Sistemas operativos.</td>
</tr>
</table>

Tabla 2 Comparación 2 de herramientas de análisis

Fuente: (González, 2015)

Una vez realizadas las respectivas comparaciones, se puede concluir que la herramienta más óptima para su implementación es la ELK, debido a que la unión de éstas 3 herramientas crea un software potente; además la Base de Datos que hace uso, puede realizar búsquedas eficientes y da respuestas rápida a las peticiones que se realizan. Sus sistema de colas, acompañado de la transformación de registos que se reliza en la propia herramienta y no en el equipo a analizar, la convierte en la herramienta perfecta para su uso, por lo no hay que configurar casi nada en el equipo que se requiere analizar. Cabe recalcar que permite manejar grandes volúmenes de datos y es compatible con todos los Sistemas Operatvos.

3.2. Implementación de herramienta seleccionada

La realización práctica del proyecto se llevó a cabo en un ambiente de software libre llamado Ubuntu, en compañía de virtualizaciones creadas con docker, el cual posee contenedores listos y completos de las herramientas que ELK hace uso. Para la implementación se debe hacer lo siguiente:

- Crear el contenedor de Elasticsearch, la cual es la Base de Datos, definiendo el puerto que éste va a usar.

```
root@bryan:/home/bryan# docker run -d -p 9200:9200 -p 9300:9300 -it -h elasticsearch --name elasticsearch elasticsearch
fcb7de86abb40d7bb57a743ebf86cc2ae499bcfd24957f66f01361985bf6c622
```

Ilustración 14 Creación del contenedor Elasticsearch

- Una vez creado el contenedor, por medio de un curl y la dirección del Elastic, se puede ver la información de éste y verificar que funciona correctamente.

```
root@bryan:/home/bryan# curl http://localhost:9200/
{
  "name" : "GClaZxb",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "q5KIjyFrR0iU_zw7kInY6w",
  "version" : {
    "number" : "5.6.9",
    "build_hash" : "877a590",
    "build_date" : "2018-04-12T16:25:14.838Z",
    "build_snapshot" : false,
    "lucene_version" : "6.6.1"
  },
  "tagline" : "You Know, for Search"
}
```

Ilustración 15 Visualización de información del contenedor Elasticsearch

- Teniendo listo la Base de Datos, se procede a instalar la herramienta kibana, enlazando o especificando que la BD será Elasticsearch; ésta es la interfaz gráfica dónde el usuario podrá visualizar los registros logs en gráficos estadísticos.

```
root@bryan:/home/bryan# docker run -d  -p 5601:5601 -h kibana --name kibana --link elasticsearch:elasticsearch kibana
31c5253acae6613d0c1826f2674c66221d8646fc8c67917da33f65b095602570
```

Ilustración 16 Instalación del contenedor Kibana

- De ingual forma por medio del comando curl y la dirección de la interfaz se puede ver datos de la herramienta.

```
root@bryan:/home/bryan# curl http://localhost:9200/_cat/indices
yellow open .kibana wAbojJ8dSJ6y_7KABBamjw 1 1 1 0 3.1kb 3.1kb
```

Ilustración 17 Visualización de información del contenedor Kibana

- Llegado a éste punto solo falta la instalación de la herramienta de recolección de datos llamada Logstash, pero antes de crear su contenedor se debe crear un archivo con extensión .conf, dónde se especificará a donde se envía los datos de salida y de se recolectan los datos entrada; además de que permite realizar algunos filtros para un análisis más personalizado.

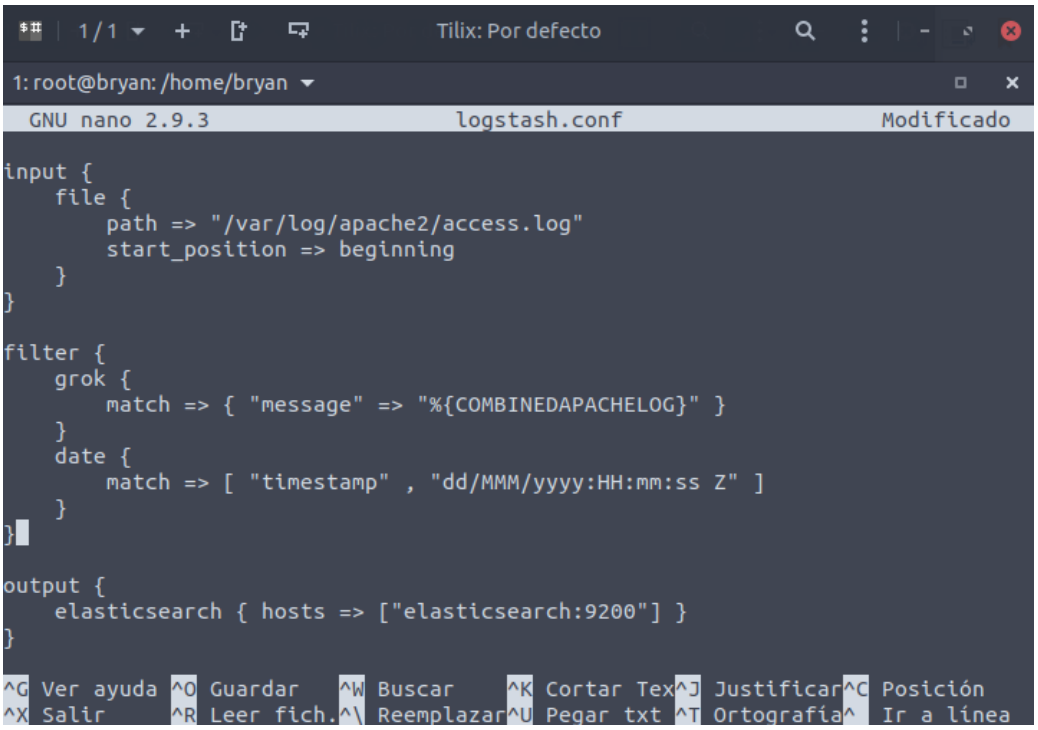


Ilustración 18 Creación y modificación de archivo de configuración del Logstash

- Una vez listo todos los pasos anteriores, se procede a instalar el logstash, enlazándolo a la Base de Datos y haciendo referencia al archivo de configuración anteriormente mencionado. Una vez realizado esto, se comenzará a crear y a estar lista para la recolección de registros.

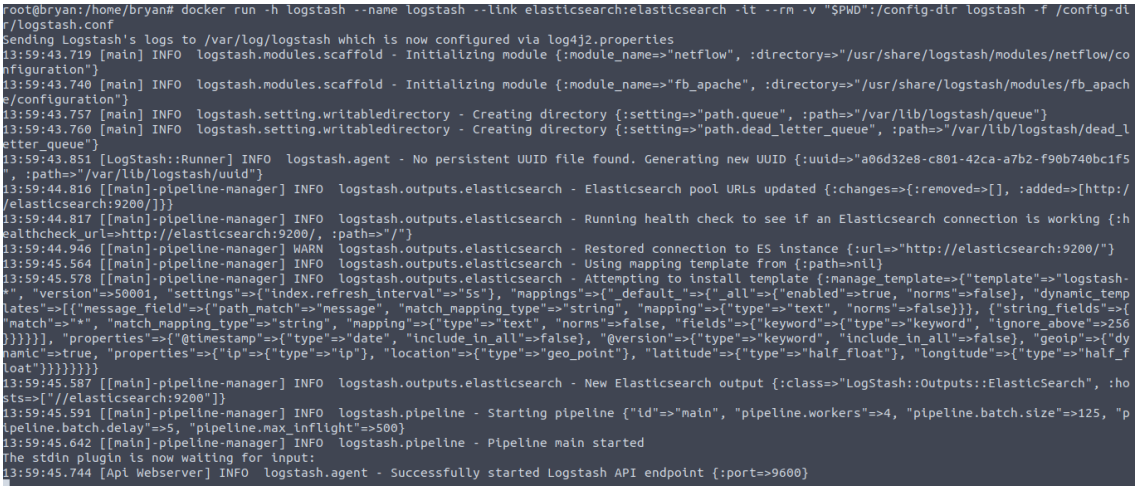


Ilustración 19 Creación del contenedor Logstash

- Por último, el usuario podrá ingresar a la interfaz por medio del navegador y ser testigo de los eventos que van apareciendo en la pantalla.

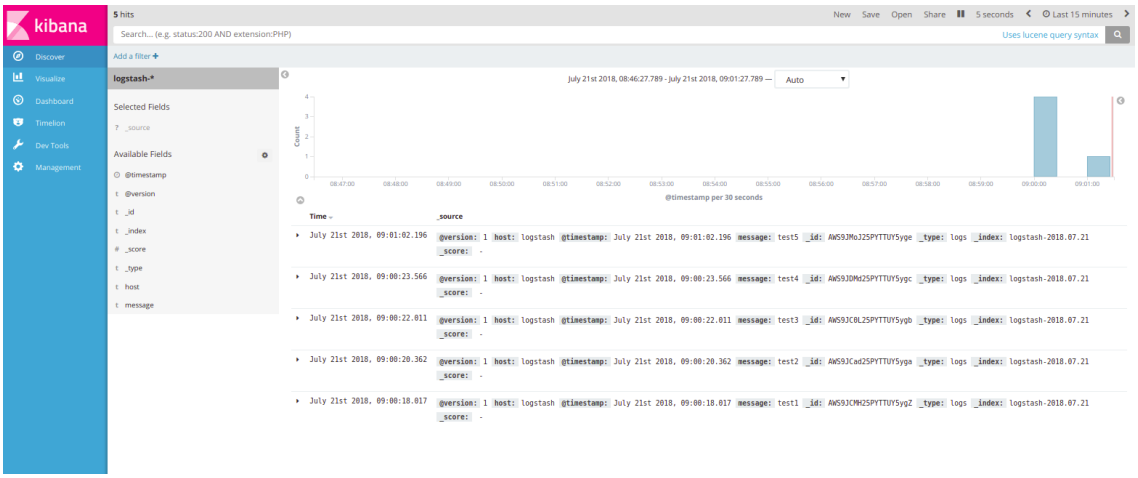


Ilustración 20 Interfaz de la herramienta ELK

3.3. Análisis de archivos logs generados

Una vez realizada la implementación de la herramienta ELK, se procedió con el envío de los diferentes archivos logs que se generaron en el servidor web que aloja el Aula Virtual de la PUCE Sede Esmeraldas, obteniendo un sinnúmero de datos que se encuentran reflejados en graficos estadísticos para un mejor entendimiento de los mismos, como se presentan a continuación:

- Como ventana principal, el usuario se encontrará con una análisis general de los archivos logs que se estan generando y enviando a la herramienta, representado por un gráfico de barras. Éste visualiza la actividad que se realiza en el servidor web que se está analizando, mostrando la cantidad de logs y el tiempo en que se crean; como se muestra en la **Ilustración 21**, la actividad del servidor web (Aula Virtual) en horas de 1:00 – 4:00 es sumamente baja, a comparación en horas de 8:00 – 12:00 donde el servidor se encuentra con una actividad un poco mayor a la anteriormente mencionada. También se puede notar que, aproximadamente en horas de 10:00 y 16:00 la actividad del servidor aumenta; aunque a diferencia de la manaña, en la tarde se hizo un mayor uso del sitio.

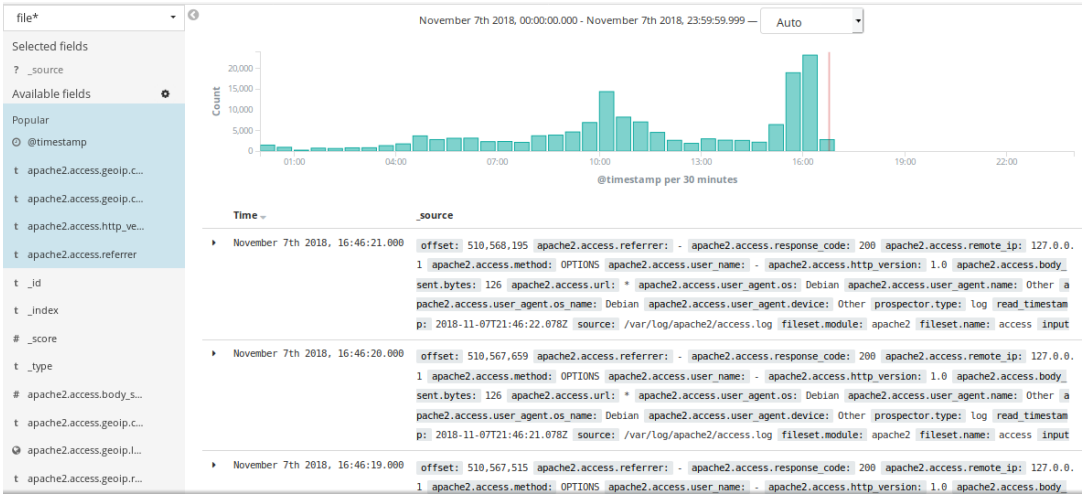


Ilustración 21 Gráfica general de la actividad del servidor web

- Los archivos logs que son recibos por la herramienta de análisis contienen cierta información relevante, como es latitud y longitud, los cuales permiten conocer la ubicación de los diferentes equipos clientes que se encuentran accediendo y haciendo uso de los servicios que brinda el servidor web que se está analizando. La herramienta ELK posee un mapa de vizualización que permite graficar las

diferentes ubicaciones que brinda ésta información, además de eso las clasifica de cierta forma qué, en el mapa coloca diferentes puntos de diferentes colores, y dependiendo de la cantidad de equipos pertenecientes a cierta region o ubicación que estén accediendo a los servicios del servidor, se graficará un punto con un color y tamaño respectivo. En la **Ilustración 22** se puede apreciar la ubicación de ciertos equipos clientes que se encuentran esparcidos alrededor de todo el mapa, tal vez de ubicaciones que la Institucón no imagina que se están realizando accesos a la página del moodle, y a su vez éstos podrían tratarse de bots que intentan romper la seguridad del sitio; mientras que en la **Ilustración 23** se encuentra más enfocada la zona de ecuador y se puede observar qué, a pesar de que hay trafíco en ciertas zonas como Esmeraldas, Guayaquil, etc; existe una gran cantidad de tráfico en Quito, ya que es el punto que cubre más área.

Ilustración 22 Visualización de ubicación de equipos clientes mapa general

Ilustración 23 Vsualización de ubicación de equipos clientes en Ecuador

- Los servidores web manejan ciertos códigos, denominados Códigos de estado HTTP; éstos consisten en describir si alguna petición que se realiza en el servidor fue completada de forma satisfactoria o no, dentro de éstos se encuentran: las respuestas realizadas con éxito, informativas, errores (clientes y servidor) y redirecciones. Ésta herramienta facilita la lectura de éstos tipos de códigos, ya que permite vizualizarlos en un gráfico de barras, como se muestra a continuación en la **Ilustración 24**, en éste se puede notar que la mayoría de las peticiones realizadas se han hecho de manera exitosa, ya que éste corresponde al código 200 perteneciente al conjunto de códigos de peticiones correctas y se encuentra resaltado de color verde; otro dato que se puede apreciar es que aproximadamente antes de las 16:00 se han realizados peticiones que no han resultado satisfactorias y que se encuentran referenciadas con el código 407 y en color rojo, ésto se debe a que éste código pertene al conjunto de código de errores y significa que una auntentificación a partir de un proxy no se está realizando de manera correcta.

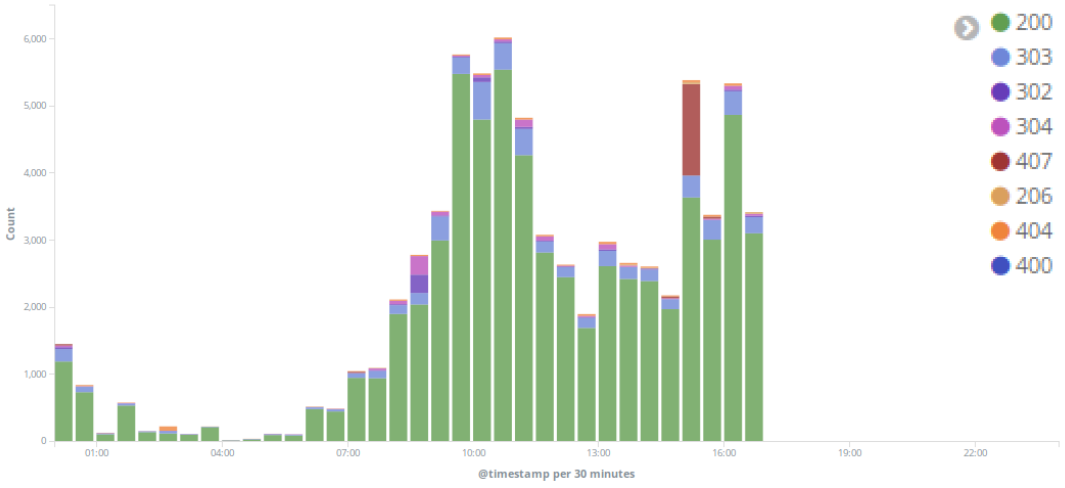


Ilustración 24 Gráfica de análisis de códigos HTTP

- Otro datos que permite analizar de forma gráfica y fácil es el Sistema Operativo con el cual los clientes estan haciendo uso para acceder a los servicios que contiene el servidor, éste se encuentra reflejado en un gráfico de pastel y clasifica los diferentes sistemas operativos con sus respectivos colores. **En la Ilustración 25** se puede notar que el gráfico indica que los sistemas operativos que son más usados por los clientes son Windows 10, que se encuentra representando por un color rojo, y android, representado por un color celeste; cabe recalcar que la

gráfica también muestra que la cantidad de uso por medio de éstos Sistemas son similares, es decir, van a la par.

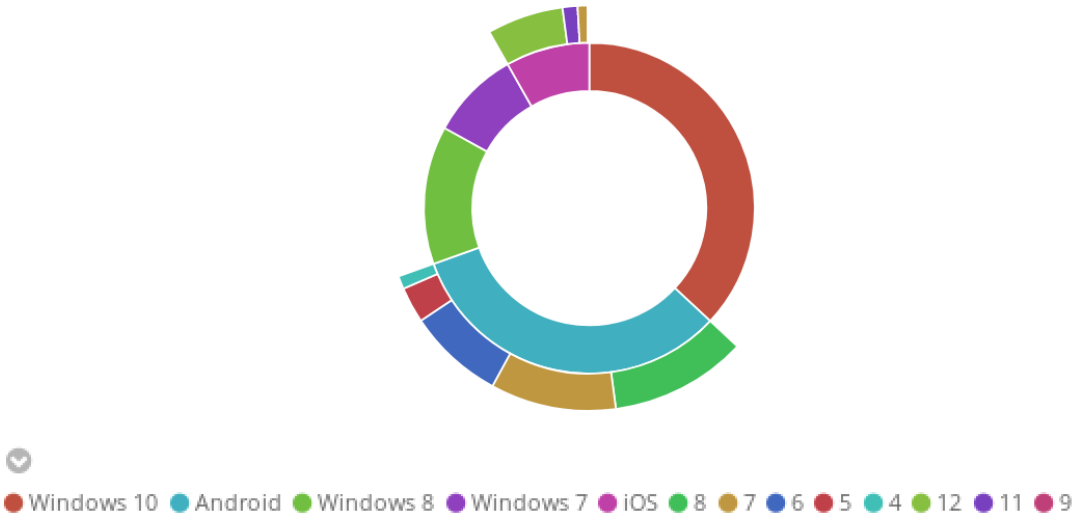


Ilustración 25 Gráfica de análisis de SO usados por equipos clientes

- En la actualidad existe una gran variedad de navegadores, los cuales son usados por los equipos clientes para acceder a los diferentes sitios web que existen en la red y al momento de ingresar a uno de éstos, ésta información es almacenada en los archivos logs del servidor. De ésta forma la herramienta ELK visualiza por medio de grafico de pastel los diferentes navegadores que son utilizados al momento de acceder al sitio web del servidor que se está analizando; como es el caso de la **Ilustración 26** dónde muestra que el navegador que más se usa para acceder al aula virtual de la PUCE Sede Esmeraldas es google chrome, tanto en ordenadores como en dispositvos móbiles.

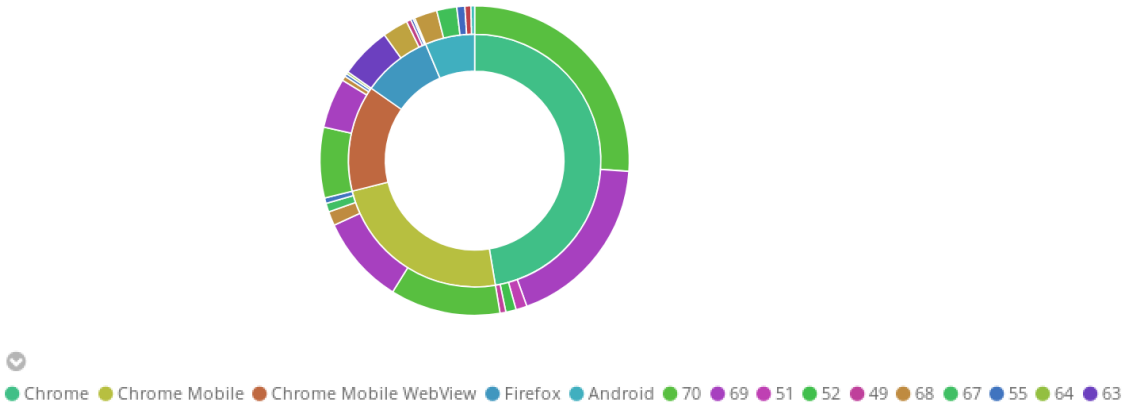


Ilustración 26 Gráfica de análisis de navegadores web usados por equipos clientes

CAPÍTULO IV

4. Discusión

Un servidor web es aquel equipo especial con conexión a internet, el cual posee un software diseñado para la transferencia de datos de hipertextos; éste se encuentra a la espera de que un usuario acceda por medio de un navegador y realice una determinada acción o petición, para responderle mediante él envió de código HTML.

Mellouk (2016) en su investigación resalta la importancia que posee los archivos logs y los eventos que se guardan en éstos, como el autor menciona, permite conocer el comportamiento, el rendimiento y detectar el acceso no autorizados al sistema alojado en el servidor web.

Con el conjunto de herramientas llamada ELK (Elasticsearch-Logstash-Kibana), la cual es una herramienta potente destinada al análisis de archivos log, se logró demostrar todos los puntos que el autor ha mencionado, ésta herramienta brinda una gran variedad de opciones que permiten facilitar y agilizar el análisis de los registros como son: la dirección IP del equipo que accede, el navegador, la ubicación que éste posee, las ubicaciones dónde se accede con más frecuencia, etc. Información que resulta sumamente útil para llevar una correcta administración de un sitio web, tomar decisiones en caso de suceder algún evento y darles solución a los problemas.

Dávila, Ortiz, y Cruz (2016) manifiestan que la mayoría de las instituciones han tomado conciencia de los riesgos que lleva la implementación y uso de los sistemas de información, y que son de suma importancia en la actualidad, esto se debe a que la mayoría de instituciones hacen uso de un sitio web para promocionar sus productos o servicios, además de que manejan el comercio electrónico, dónde el usuario por medio de ésta puede pagar los productos que desee obtener; en consecuencia a esto se han convertido en uno de los puntos más atractivo de los delincuentes informáticos y una falla operativa puede traerles pérdidas muy grandes.

ELK brinda una gran flexibilidad al momento de realizar el análisis, ya que permite seleccionar los puntos que se quieren analizar, de modo a que brinda una información detallada de los eventos que sucenden en el servidor; permite seleccionar el gráfico estadístico que mas se adapte a las circunstancias; y el momento en el que se genere cierto error de operatividad, inmediatamente se verá reflejado en el analizador, permitiendo

darle solución de una forma rápida; además de permitir ver dónde y cuando un supuesto usuario ingresa al sistema; de ésta forma se podrá llevar un correcto control para no sufrir pérdidas.

Gómez, Candela, y Sepúlveda (2013) resaltan que la informática ha brindado muchos beneficios a la sociedad, permitiendo simplificar y agilizar múltiples procesos que comúnmente las personas realizan en su vida cotidiana; pero así como ha generado beneficios, también ha creado nuevos peligros en lo que respecta la seguridad de la información. Esto se origina ha partir de la falta de control y de monitoreo que muchas instituciones dejan pasar por alto, quedando expuestas a robo de información, suplantación de identidad, etc.

De aquí un gran aporte de la herramienta analizadora ELK, ya que permite llevar un control y monitoreo de las sesiones que se realicen en un determinado sistema, de ésta forma se puede conocer ¿cuando? y ¿a qué? entró un usuario al sistema; llevando un control de sesiones y permitiendo minizar posibles pérdidas de información

Ferreira (2015) afirma que los registros logs, además de permitir llevar el control de un sitio web, también brindan otro beneficios en lo que respecta el análisis; dónde no solo se enfoca a la seguridad, sino que permiten conocer ciertos aspectos como: qué página o qué servicio los usuarios utlizan más frecuentemente, el rendimiento o tiempo de ejecutarse cierto servicio, etc; en caso de realizarse en una empresa dedicada al E-commerce, el analizador sería de mucha utilidad, ya que permitiría conocer cual de sus productos es que más se vende y en qué país es su fuente más alta de ingresos, obteniendo un gran beneficio de marketing beneficiándose la empresa.

Finalmente se puede decir que un analizador de logs ofrece múltiples ventajas, ya que permite llevar un monitoreo completo del sistema que se analice, conocer los errores que ocurren dentro de un sistema, al igual que cada una de las acciones que se realicen en éstas; y con la ayuda de ELK es mucho más facil, debido a que es un sistema multi plataforma y compatible con varios métodos de entrada para la obtencion de archivos logs, al igual de como se muestra en los resultados su implementación es fácil y de Software Libre, por lo que no se necesitan pagar por su instalación, además de que su comunidad se encuentra sumente activa.

CAPÍTULO V

5. Conclusiones y Recomendaciones

5.2. Conclusiones

Realizada la parte práctica de la investigación y partiendo de los resultados se puede concluir que:

- Los registros logs, son archivos de gran importancia en lo que respecta el análisis y monitoreo de un sistema; debido a que éstos almacenan las diferentes actividades que se realizan en éste, brindándo información relevante cómo el tiempo, ubicación, origen de la acción, quien accede al sistema, etc.
- En la actualidad existen una gran variedad de herramientas que permitan realizar el análisis de archivos logs, pero las que más se destacan en lo que respecta a software libre son Graylog y ELK, debido a la gran potencia que éstas brindan al momento de llevar a cabo un determinado análisis de éste tipo de archivos; pero a pesar de qué éstas herramientas muestran características similares, su diferencia es abismal; debido a que la mayoría de las funcionalidades que ofrece graylog se debe a la implementación de módulos, a diferencia de que ELK ya posee herramientas que permiten obtener un análisis completo, y al igual que graylog permite adaptar módulos para realizar análisis más avanzados.
- La implementación de la herramienta ELK, con ayuda de los contenedores virtuales Docker, resulta muy útil; debido a que éstos contenedores ya cuentan con todas las herramientas y extensiones listas para su uso.
- ELK resultó ser una herramienta sumamente útil y de gran importancia para un correcto control del servidor web que aloja el Aula Virtual de la PUCE Sede Esmeraldas, debido a que brinda una información detallada de las diferentes actividades que se han realizado en éste, como es mostrar la ubicación de los equipos clientes que acceden a los servicios que brinda el servidor anteriormente mencionado, como tambien los códigos HTTP y los errores que se van originando; además de que presenta la información en una forma muy fácil de interpretar y comprender, como son los gráficos estadísticos. Por último, la herramienta aporta una gran flexibilidad al momento de modificar, personalizar y crear los diferentes paneles o gráficas que la Institución requiera para un mejor control del mismo.

5.3. Recomendaciones

- El personal que administrará el Sistema ELK en el Departamento de TIC de la PUCE Sede Esmeraldas, debe ser capacitado para que puedan comprender y aprender el funcionamiento de la herramienta y poder sacarle provecho.
- El personal que estará a cargo de ELK debe ser capacitado sobre como trabajar con archivos logs, la información que guardan y que se puede hacer con ésta; de esta forma se podrá entender mucho mejor el funcionamiento de la herramienta de análisis y poder realizar búsquedas más avanzadas.
- Debido qué, la página de la PUCE Sede Esmeraldas se encuentra constantemente en uso por los diferentes clientes que acceden a ésta, la información que almacenan los archivos logs es sumamente grande; por ende el Depeartamento de TIC debe posser equipos adecuados para el correcto funcionamiento de la herramienta, permitan almacenar una cantidad masiva de datos.
- Como en los resultados se muestra, existen accesos de diferentes partes del mundo, tal vez lugares que ni la misma Institución pensaría que pueden estar intentando acceder al sitio web de la PUCE Sede Esmeraldas, por ende el encargado de administrar la herramienta ELK debe estar en constante retroalimentación de las diferentes vulnerabilidades y ataques que se pueden generar hacia un servidor web, de ésta forma se con la ayuda de la herramienta se podrá minizar los diferentes riegos que existan y proteger de mejor forma los sistemas de la Institución.

CAPÍTULO VI

6. Referencias

6.1. Glosario

TIC: Conjunto de tecnologías desarrolladas para una información y comunicación más eficiente.

XML: Metalenguaje extensible de etiquetas que fue desarrollado por el Word Wide Web Consortium.

HTTP: Se trata de un protocolo de comunicación que posibilita la circulación de información a través de la World Wide Web (WWW).

OSI: Modelo creado por la ISO para la interconexión en un contexto de sistemas abiertos.

PHP: Lenguaje de código abierto muy popular especialmente adecuado para el desarrollo web.

JSP: Tecnología orientada a crear páginas web con programación en Java.

HTML: Lenguaje de marcado que se utiliza para el desarrollo de páginas de Internet.

UDP: Protocolo sin conexión que, como TCP, funciona en redes IP.

TCP: Protocolo que permite a dos anfitriones establecer una conexión e intercambiar datos.

.NET: Componente de software que puede ser o es incluido en los sistemas operativos Microsoft Windows.

IIS: Conjunto de servicios para servidores usando Microsoft Windows.

ELK: Herramienta de administración de registros.

BD: Base de datos.

IP: número que identifica de forma única a una interfaz en red de cualquier dispositivo.

FTP: Protocolo de Transferencia de Archivos.

6.2. Referencias

Abreu, J. (2014). El Método de la Investigación. *International Journal of Good Conscience*, *9*(3), 195–204. Retrieved from http://rdigital.unicv.edu.cv/bitstream/123456789/106/3/Libro metodologia investigacion este.pdf

Avella, J., Calderón, L., & Mateus, C. (2015). *Guía metodológica para la gestión centralizada de registros de seguridad a través de un SIEM*. Botoga. Retrieved from http://hdl.handle.net/10983/2847

Behar, D. (2010). Introducción a la Metodología de la Investigación. *Shalom*, *1*(978-959-212-783–7), 1–94. https://doi.org/10.1017/CBO9781107415324.004

Berzal, F., Cortijo, F., & Cubero, J. (2010). *Desarrollo Profesional de Aplicaciones Web con ASP.NET*. Retrieved from http://elvex.ugr.es/decsai/csharp/pdf/web/web-book-a4.pdf

Carrión, B. (2015). Diseño e Implementación de una solución de gestion centralizada de logs de aplicaciones, sistemas y dispositivos basada en Logstash que permita la creación de cuadros de mando para explorar, analizar y monitorear eventos de seguridad, 67. Retrieved from http://openaccess.uoc.edu/webapps/o2/handle/10609/42250%5Cnhttp://hdl.handle.net/10609/42250

Chung, K. (2013). JavaServer Pages™ Specification. Retrieved from http://download.oracle.com/otn-pub/jcp/jsp-2_3-mrel2-eval-spec/JSP2.3MR.pdf?AuthParam=1510008178_fb0df993f3a741f92748f78a1ac81b89

Dávila, D., Galvis, A., & Vivas, R. (2015). Sitio Web Como Estrategia De Enseñanza En La Educación Para La Sostenibilidad. *Praxis & Saber*, 115–138. Retrieved from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S2216-01592015000100006&lang=pt

Dávila, G., Ortiz, F., & Cruz, F. (2016). Cálculo del valor en riesgo operacional mediante redes bayesianas para una empresa financiera. *Contaduria y*

Administracion, *61*(1), 176–201. https://doi.org/10.1016/j.cya.2015.09.009

De la Fuente, T. (2012). OSI, la pila teórica de protocolos de red, 22. Retrieved from https://blyx.com/public/docs/pila_OSI.pdf

Ferreira, R. (2015). A análise de logs como estratégia para a realização da garantia do usuário. *Em Questão*, *21*(n. 3), 150–170. https://doi.org/http://dx.doi.org/10.19132/1808-5245213.150-170

Gómez, C., Candela, C., & Sepúlveda, L. (2013). Seguridad en la configuración del Servidor Web Apache, *9*, 31–38.

González, A. (2015). Propuesta de Arquitectura Distribuida para la gestión de Logs. Retrieved from https://e-archivo.uc3m.es/bitstream/handle/10016/22278/PFC_Abel_Cal_González.pdf

Granados, R. (2014). *Desarrollo de aplicaciones web en el entorno servidor (UF1844)*. ic editorial. Retrieved from http://www.tierradelazaro.com/wp-content/uploads/2016/08/UF1844-Desarrollo-de-aplicaciones-web-en-el-entorno-servidor.-IFCD0210-.pdf

Guamán, R. (2011). Seguridad en Entornos Web para Sistemas de Gestión Académica, 1–47. Retrieved from http://repositorio.educacionsuperior.gob.ec/bitstream/28000/120/1/Seguridad de entornos web.pdf

Hernández, A., & Mejia, J. (2015). Guía de ataques, vulnerabilidades, técnicas y herramientas para aplicaciones web. *ReCIBE*, *4*(1), 17. Retrieved from http://recibe.cucei.udg.mx/revista/es/vol4-no1/pdf/computacion05.pdf

Hernández, A., & Porven, J. (2016). Procedimiento para la seguridad del proceso de despliegue de aplicaciones web. *Revista Cubana de Ciencias Informáticas*, *10*(2), 42–56. Retrieved from http://rcci.uci.cu

IPA. (2011). How to Secure Your Website. *It Security Center (Isec) Information-Technology Promotion Agency*, (April), 105. Retrieved from papers3://publication/uuid/C8F38793-41EA-4007-8C15-1F9780E69973

Labrador, R. (2010). Administración De Servidores Linux (Ubuntu/Fedora). Retrieved from

https://www.informatica.us.es/~ramon/articulos/AdminLinuxUbuntuFedora.pdf

Langhnoja, S., Barot, M., & Mehta, D. (2012). Pre-Processing: Procedure on Web Log File for Web Usage Mining. *International Journal of Emerging Technology and Advanced Engineering ISO Certified Journal*, *2*(12), 419–423. Retrieved from www.ijetae.com

Mejía, M. (2014). Diseño e implementación de una estrategia de seguridad de la información, 1–25.

Mellouk, M. (2016). Diseño e implementación de un sistema para la recogida de logs en sistemas distribuidos. Retrieved from https://www.uam.es/ss/Satellite/es/1242647861998/contenidoFinal/Grupos_de_Investigacion.htm

Metaute, A. (2013). Seguridad en Aplicativos Web. Retrieved from http://diposit.ub.edu/dspace/bitstream/2445/49106/1/AdrianHermoso_memoria.pdf

Mitchell, C. (2011). Securing websites, (October). Retrieved from https://www.sophos.com/en-us.aspx

Monte, A. (2016). Diseño e implementación de infraestructura NIDS (Network Intrussion Detection System) para PIMES. Retrieved from https://riunet.upv.es/handle/10251/88856

Murugiah, K. (2006). Guide to Computer Security Log Management. *National Institute of Standards and Technology*, 1–72. https://doi.org/10.6028/NIST.SP.800-92

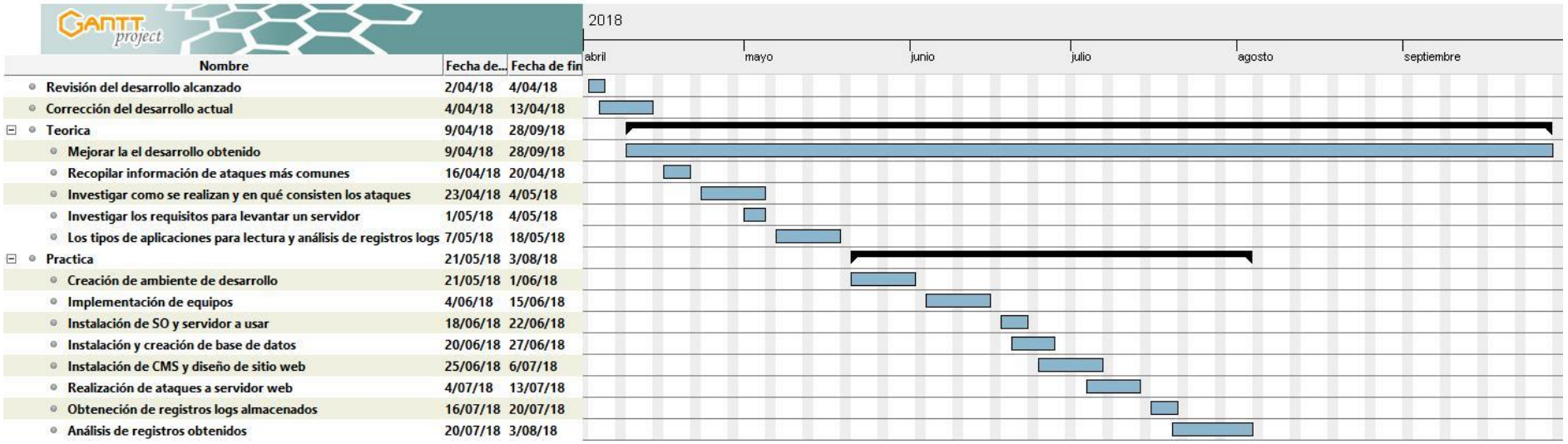
Palomo, M. (2013). Programación en PHP a través de ejemplos, 1–54. Retrieved from http://servicio.uca.es/softwarelibre/publicaciones/apuntes_php

Rubinos, A., & Nuevo, H. (2011). Seguridad en bases de datos Security Database. *Revista Cubana de Ciencias Informáticas (RCCI)*, *5*(Sistema de bases de datos), 16. Retrieved from ehttp://www.redalyc.org/articulo.oa?id=378343671005

Tulloch, M. (2013). *Introducing Windows Server 2012 R2* (Vol. 15). Retrieved from https://books.google.com/books?id=IlxuAwAAQBAJ&pgis=1

6.3.ANEXOS

6.3.1. ANEXO 1



Nombre	Fecha de...	Fecha de fin
Revisión del desarrollo alcanzado	2/04/18	4/04/18
Corrección del desarrollo actual	4/04/18	13/04/18
Teorica	9/04/18	28/09/18
Mejorar la el desarrollo obtenido	9/04/18	28/09/18
Recopilar información de ataques más comunes	16/04/18	20/04/18
Investigar como se realizan y en qué consisten los ataques	23/04/18	4/05/18
Investigar los requisitos para levantar un servidor	1/05/18	4/05/18
Los tipos de aplicaciones para lectura y análisis de registros logs	7/05/18	18/05/18
Practica	21/05/18	3/08/18
Creación de ambiente de desarrollo	21/05/18	1/06/18
Implementación de equipos	4/06/18	15/06/18
Instalación de SO y servidor a usar	18/06/18	22/06/18
Instalación y creación de base de datos	20/06/18	27/06/18
Instalación de CMS y diseño de sitio web	25/06/18	6/07/18
Realización de ataques a servidor web	4/07/18	13/07/18
Obteneción de registros logs almacenados	16/07/18	20/07/18
Análisis de registros obtenidos	20/07/18	3/08/18

Anexo 1. Cronograma de actividades

6.3.2. ANEXO 2

Servidores web:

- ¿Qué son?
- Funcionamiento
- Herramientas que lo conforman
- Plataforma
- Arquitectura
- Modelos de capas
- Ventajas y desventajas

Seguridad:

- Introducción
- Amanezas
- Ataques
- Mecanismo de seguridad
- Tipos de ataques
- Características
- Tipos de ataques
- Principios básicos

Logs:

- Funcionamiento
- Por qué son necesarios?
- Dificultad que presentan
- Arquitectura
- Como llevarlo a cabo
- Que almacenan
- Control y monitoreo

Anexo 2. Lista base de aspectos a investigar