Visualización de Tráfico de Red con Parallel Node-Links

Juan Alvarez Piedra, Osael Josue Jimenez Murillo y
Pablo Andrés Rodríguez Blanco

Escuela de Ingeniería,
Universidad Latinoamericana de Ciencia y Tecnología,
ULACIT, Urbanización Tournón, 10235-1000
San José, Costa Rica
jalvarezp944@ulacit.ed.cr,ojimenezm724@ulacit.ed.cr,
prodriguezb047@ulacit.ed.cr
http://www.ulacit.ac.cr

Resumen Hoy la cantidad de información que circula en la red es vasta. Dentro de todo este tráfico, existen elementos que han sido introducidos con la intención de causar daño, ya sea para robar información o borrarla. Existen programas que recopilan información sobre dichos elementos y la guardan en archivos históricos. El gran volumen de estos archivos genera la necesidad de buscar herramientas que permitan su lectura y análisis rápido a fin de definir patrones para mitigar el riesgo de daños futuros. Estas herramientas se constituyen en la respuesta a la interrogante ¿cómo se puede leer gran cantidad de información, procesarla y mostrarla de tal forma que el usuario la comprenda fácilmente y pueda analizarla? La visualización -proceso para presentar la información- se da como una solución, ofrece herramientas que condensan altos volúmenes de datos en gráficos que permiten leer y comprender su significado. Lo anterior implica pasar por una serie de pasos que van desde extracción de los datos, hasta su análisis, filtrado y combinado con tecnologías de representación que generen un modelo totalmente descriptivo y dinámico. El presente proyecto se basa en trabajos relacionados con la visualización de la información y consiste en el desarrollo de una herramienta que permite analizar datos dañinos históricos y reflejar los resultados de ese análisis de una manera dinámica y amigable.

Keywords: Visualización, Gráficos, Internet

1. Introducción

Gran parte de la información digital que se maneja a nivel mundial viaja a través de Internet haciendo uso de múltiples dispositivos como servidores, PCs, conmutadores, enrutadores, entre otros. El envío de esta información produce una extraordinaria cantidad de registros o archivos de bitácora, los cuales con-

tienen las trazas del recorrido desde su origen hasta su destino, que incluyen su IP[1] origen, su IP destino, direcciones de acceso al medio[2] y los puertos utilizados.

Analizar estos archivos de manera manual es un ejercicio tedioso y complicado, dada la enorme cantidad de información que estos contienen, la cual es generada de manera automática por los dispositivos.

A lo largo de la historia, el ser humano se ha encontrado ante el reto de analizar gran cantidad de información. El uso de mapas y otros formatos para mostrar la información en las áreas de la geografía y las ciencias sociales, así como los gráficos para el resumen y la representación de datos en las matemáticas, estadística, y contabilidad, son claros ejemplos de instrumentos utilizados para dichos análisis. En ese mismo sentido, en este proyecto se propone la visualización como medio para la lectura de los registros dañinos y su análisis ágil. Esto vendría a responder a la pregunta de ¿cómo se puede leer gran cantidad de información, procesarla y mostrarla de tal forma que el usuario la comprenda fácilmente y pueda analizarla?

Lo anterior repercutiría en eficiencia y rapidez de la toma de decisiones en cuanto a la seguridad informática. Por ejemplo, ante la detección de un ataque a nivel de red, el usuario puede tomar medidas y acciones preventivas y/o correctivas, dado que los ataques más comunes se pueden presentar en resúmenes mostrados en gráficos.

El objetivo del proyecto es desarrollar una herramienta web que permita representar de una forma gráfica, la información de las trazas dañinas, recopilada en los archivos producidos por el Snort. Para estos efectos se utilizará el gráfico de barras, no así otro tipo de gráficos, mapas o formatos. Para la puesta en prueba de la herramienta se cuenta con un archivo obtenido de un Snort, por lo que no es necesaria la recopilación de información adicional.

Para el presente trabajo se realizó una investigación bibliográfica con el fin de asentar la base teórica sobre el tema. Posteriormente, se desarrolló una herramienta de visualización, para la cual se utilizaron las tecnologías de MEAN[3] por su robustez en el desarrollo de aplicaciones web. El proceso del proyecto se documenta en este trabajo y se concluye con el análisis de los resultados obtenidos.

1.1. Metodología

La investigación bibliográfica se enfoca en el estudio de los temas como 'visualización de la información' y 'seguridad de la información' con el fin de obtener conocimientos que ayuden al desarrollo del trabajo. El estudio la documentación se centra en encontrar información sobre la problemática de la visualización, sobre herramientas relacionadas y sobre resultados de investigaciones afines, todo

[1] Etiqueta numérica que identifica un dispositivo utilizando el protocolo IP (Internet Protocol

[2] Conocidas en inglés como direcciones MAC por Media Access Control.

[3] Tecnologías que reciben dicho acrónimo debido a sus iniciales: Mongo, Express, AngularJS y NodeJS.

ello con el fin de contar con un soporte teórico que ayude en el desarrollo de la herramienta por elaborar.

2. Estado del arte

La cantidad de datos que se trasmite a nivel mundial crece constantemente de una manera exorbitante. Los sistemas presentes para hacer uso de la gestión de redes, han dado un paso adelante con el uso de la visualización, que ayuda a entender la información que se transmite por las redes. Adentrarse en dicha área permite la representación de miles de datos en imágenes que son más fáciles de comprender, hace sencillo poder apreciar cambios, patrones, percances de seguridad y de desempeño (Bhardwaj y Singh, 2015).

Esto lleva a la pregunta ¿ qué es visualización? La visualización de la información es un mapeo de los datos en una representación gráfica (Manovich, 2010). La incursión en esta área facilita el análisis de los datos ya que se puede observar de una manera general en la representación gráfica toda la información que se está mapeando. Esto permite realizar análisis de grandes cantidades de información, la cual se puede relacionar entre sí o manejar en forma independiente a través de gráficas, análisis temporales e incluso análisis geo posicionales.

El desarrollo de una visualización requiere de una serie de pasos(Fry, 2016) que se listan a continuación:

- Adquirir: Extraer de la fuente o fuentes la información.
- Analizar: Proceso donde se crea una estructura a los datos, ya que estos se almacenan de forma masiva.
- Filtrar: En este procedimiento es donde se conservan los datos con contenido relevante o necesario.
- Minería: Se procede con la aplicación de métodos matemáticos con el objetivo de que la información sea exacta.
- Representación: Empleo de gráficos, ya sea de barras o lineal, pictogramas y mapas.
- Refinado: Adición de tecnologías como HTML5[4], CSS3[5] y SVG[6] para que la representación sea más amigable para el usuario o interactiva.
- Interacción: Corresponde a la utilización de manipulación de datos y controladores a través de las tecnologias mencionadas.

La fuente más grande de información pública que existe en la actualidad es el Internet, no solo por la gran cantidad de información que se encuentra almacenada allí, sino también porque va creando nueva información a cada instante y desde todo lugar.

[4] Lenguaje de programación que recibe dicho acrónimo debido a sus iniciales: HyperText Markup Language, el número cinco es la versión actual.

[5] Lenguaje de programación que recibe dicho acrónimo debido a sus iniciales: cascading style sheets, el número tres es la versión actual.JS y NodeJS.

[6] Lenguaje de programación que recibe dicho acrónimo debido a sus iniciales: Scalable Vector Graphics.

Snort es una herramienta que permite captar y detectar los ataques que se están realizando en una red especifica (Roesch y cols., 1999). Estos datos son guardados en archivos. Más adelante se pueden utilizar para investigar incidentes de seguridad y así determinar si un ataque fue llevado a cabo y cómo poder tomar las acciones correspondientes. Sin embargo, cuando se desea realizar un análisis más a fondo, como ¿ cuáles son los ataques más comunes? o ¿ en qué periodo se registra la mayor cantidad de ataques?, esta información puede ser muy grande para un análisis de forma tradicional. Es aquí cuando la visualización toma un papel importante, debido que a través de ella se puede observar todos los registros que la herramienta generó de una manera gráfica, y así realizar el análisis de manera más sencilla y encontrar patrones o riesgos de forma más rápida, que trabajando sobre los registros de la información.

Con el fin de obtener toda la información de manera sencilla, en el proyecto mencionado, se puede apreciar un conjunto de herramientas que realizan un análisis tanto de forma general como específica de los componentes de una red o sistema. Esto hace pensar que la calidad y cantidad de datos que hay en una red es bastante extensa como para ir traza por traza evaluando qué tan exacta es la información encontrada. Para lo anterior, Janowski propuso una serie de ecuaciones matemáticas que proporciona datos más precisos en diferentes tipos de ataques como lo es la "Denegación de servicios" donde la cantidad de datos aumenta de una forma constante o que permite que se desarrolle a través de los algoritmos para ver el comportamiento que ocurrirá en ciertas situaciones (Janowski, 2010).

Como se ha mencionado anteriormente, ver grandes volúmenes de información en forma de texto puede que no llame la atención del lector o pueda que no se entienda su fin. Verbet demostró que el uso de tableros con datos y contenido didáctico hizo que a los estudiantes les fuera más fácil ver el material de clases de una forma más directa en contraparte a su estudio mediante la lectura del texto(Verbert y cols., 2014).

Hoy, los tiempos de respuesta son cruciales si se quiere tener medidas correctivas eficientes. El uso de la información es lo primero que se debe analizar y así reaccionar oportunamente sin depender de la situación.

La visualización es un área con muchas técnicas tanto para presentar la información como para analizar los registros. Explican Sushilkumar y Nirkh qué técnicas de visualización y de detección de anomalías pueden ser empleadas para realizar análisis forense digital para diferentes fines, como detección de anomalías y fraude(Chavhan y Nirkhi, 2012).

Es interesante como un área conlleva a la creación de nuevas técnicas y tecnologías. Por ejemplo, el uso de minería de datos, reglas de asociación o tipos de investigación, a través de los cuales se proporciona un filtrado de la información con la que se obtiene valores relevantes tanto del momento actual como históricos. Las técnicas gráficas como uso de barras, gráficos circulares y gráficos en tres dimensiones, conceden una mejor apreciación y hasta podrían crear una impresión diferente en la persona que les dé uso.

Otro punto de vista que no se ha mencionado hasta el momento es aquel donde se evalúa los valores que son muy comunes y que llegan a tener un comportamiento atípico, para lo cual hay que prestar atención a las representaciones que quedan aparte de la mayoría de la gráfica.

En concordancia con lo anterior, la herramienta SQC que analiza el tráfico de Internet es capaz de clasificar ataques e incluso detectar las más pequeñas anomalías(Beom-Hwan y Yoon, 2011). Después de ver los resultados de esta investigación es interesante apreciar cómo de una forma intuitiva es posible detectar si la red está bajo ataque, ya que toma en cuenta una gran cantidad de variables en tiempo real que hace posible saber su estado actual.

Se ha mencionado un poco sobre ambientes con altos volúmenes de información, así como trabajar con incidentes aislados, pero hay una tercera forma de realizar estos estudios, y es por medio de la implementación de un ambiente para la detección de intrusos, ya que puede contribuir a un grado de protección contra los posibles ataques de red donde se inspecciona y administra el sistema de detección de intrusos (Ennert, Mados, y Dudláková, 2014). Después de contemplar soluciones semejantes y ambientes de trabajo, hay que comprender que para realizar todo esto, hay que llevar a cabo un marco de trabajo con diferentes etapas que se dirijan hacía una mejor detección e interpretación de los resultados. Así lo plantea Andy Luse (2008) con el empleo de diferentes tecnologías, ya sean de código abierto o cerrado, para apoyarse en cada paso de captura de datos, filtrado de la información, relaciones o la extracción (Luse, 2008).

3. Desarrollo y análisis de resultados

Antes de hablar sobre los usos de la herramienta, se desea introducir ciertos términos que se utilizan. En la tabla 1 se presentan los nombres de los componentes usados para desarrollar la aplicación, así como las diferentes técnicas implementadas a través de su construcción.

Se usa MEAN[7] debido a que es una tecnología moderna que ha demostrado tener ciertas ventajas sobre lo ya conocido con respecto a la manera de crear aplicaciones web y su funcionamiento.

Además, se percibe como una ventaja que la mayoría del lenguaje para desarrollar la herramienta es JavaScript, el cual se adapta al uso de eventos y el modelo MVC[8], que la tecnología Express permite emplear. La base de datos MongoDB concede recabar una gran cantidad de registros rápidamente sin afectar los tiempos de respuesta y además la estructura de los datos puede variar entre los registros; por lo tanto, brinda la flexibilidad de ir agregando nuevas variables a los datos obtenidos.

Al iniciar la aplicación se ejecuta un comando que importa los datos recolectados del Snort, los cuales han sido previamente almacenados en la base de

[7] Tecnologías que reciben dicho acrónimo debido a sus iniciales: Mongo, Express, AngularJS y NodeJS.

[8] Patrón de arquitectura de softeware, dicho acrónimo debido a sus iniciales: Model-View-Controller

Término	Función
MongoDB	Base de datos no relacional que almacena los registros en documentos en formato JSON
Express	Marco de trabajo que funciona en conjunto a Node.js
AngularJS	Marco de trabajo para desarrollar la capa de presentación
Node.js	Marco de trabajo basado en el motor V8 JavaScript
JSON	Formato ligero de intercambio de datos
JavaScript	Lenguaje de programación que se ejecuta desde el lado del buscador, permite crear diferentes efectos de forma dinámica
Estructura de datos	Grupo de datos con características similares
Base de datos no relacionales	Bases de datos que no usan estructuras estáticas y son versátiles a la hora que extraer la información
Protocolo	Reglas que permiten a dos dispositivos comunicarse entre si

Tabla 1. Tabla de términos.

datos. El usuario podrá ver tanto los diferentes valores que ya se han mencionado como las IP de origen y destino, fechas, horas, tipos de ataques, entre otros. El usuario será capaz de seleccionar uno de los valores y observar cómo se asocia con las demás variables en la gráfica. Esto se logra gracias a la utilización de estructuras de datos que permiten elaborar comparaciones de los valores y vincularlos. Cada barra representa una variable obtenida desde la base de datos, en la cual los diferentes valores se agrupan en secciones siendo grandes o pequeños dependiendo de la cantidad de veces que aparezca el valor en los registros. Las relaciones de estos valores se distinguen a través de las líneas que se enlazan a cada barra. Así, ya sea un operario, un gerente informático o una junta de gerentes serán capaces de comprender los datos que se muestran. Por ejemplo, al obtener los resultados finales, el gerente del área de seguridad informática podría identificar que al termino de cada mes durante la actualización de los sistemas operativos, el tráfico de red se incrementa a través de protocolos pero no se acciona ningún fallo de seguridad. A la siguiente actualización de sistemas operativos, la red es obstruida con una sobrecarga en el tránsito de red contra objetivos específicos, como los servidores donde se almacena la información confidencial. Este comportamiento se puede apreciar como una constante a través de los meses.

Por lo tanto, el gerente de seguridad de informática puede exponer ante su equipo que existe una mala configuración durante los procesos de actualización y

que los atacantes llevan a cabo primero un escaneo para identificar las vulnerabilidades para después poder aprovecharlas y realizar ataques específicos. Gracias a esto, las correcciones se elaborarán de manera más enfocada a las debilidades existentes y no se hará una inversión de tiempo y de esfuerzo para poder dar con una solución.

Como se ha mencionado, estas representaciones pueden ser analizadas por personal con poco o ningún conocimiento técnico, así como por la gerencia, quien podría comprender el significado y proceder con la toma de decisiones; por ejemplo, si para poder aplicar acciones correctivas se debe invertir en equipo o capacitaciones, la información que se obtiene de los resultados de las gráficas, puede justificar la inversión.

3.1. Ejemplo de uso

A continuación se presentan las imágenes de la aplicación en ejecución. En la Figura 1 se observa que en primera instancia existen diez columnas con todos los valores obtenidos del archivo Snort, donde todas las relaciones son mostradas desde primer vértice hasta el último.

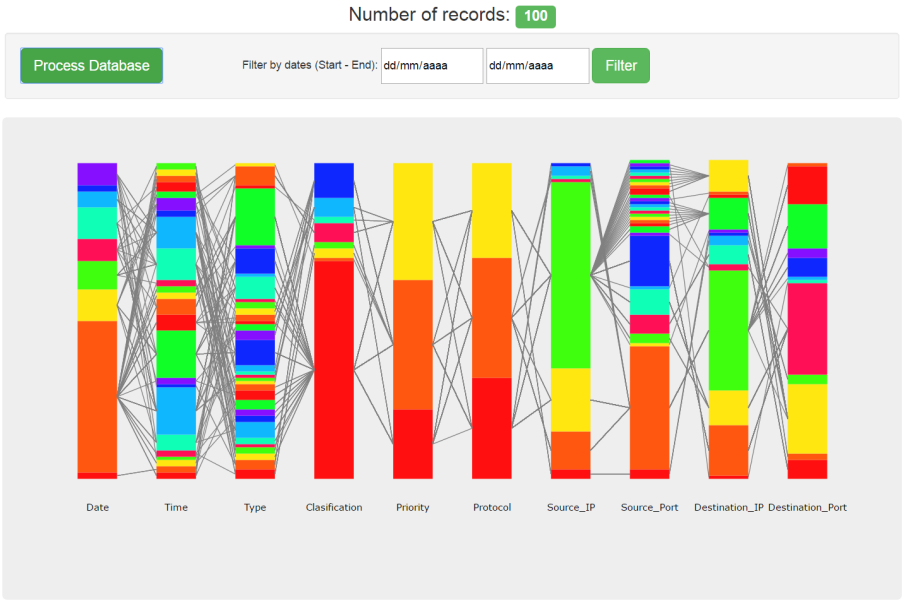

Figura 1. Datos cargados en primera instancia

Los valores asociados son: las fechas, horas, tipos, clasificación, prioridad, protocolo, IP de origen, puerto de origen, IP de destino y puerto de destino. El usuario final podrá ver que a primera vista algunos valores son constantes y

se entrelazan con otros casi formando un patrón pero aún no es una vista más granular.

En la Figura 2, la vista brinda más información, ya que se elige una sección de una de las barras y se compara con una sección de otra barra. Como se puede observar, las líneas presentes se muestran en tres tipos de colores diferentes, a saber; grises en las relaciones de los registros que no contienen los valores seleccionados, naranja para las líneas que tienen relación de los registros que contienen el valor seleccionado y rojo para los registros que contienen tanto el valor seleccionado como el que sirve de base de comparación. Al seleccionar las diferentes variables que hay en cada columna, se podrá ver el valor, la cantidad de veces que se encuentra y el porcentaje con respecto al total de los registros. En conclusión, es posible analizar cada aspecto que esté relacionado con valores específicos, ya sea el tiempo, fechas, los IP de origen o los IP destino.

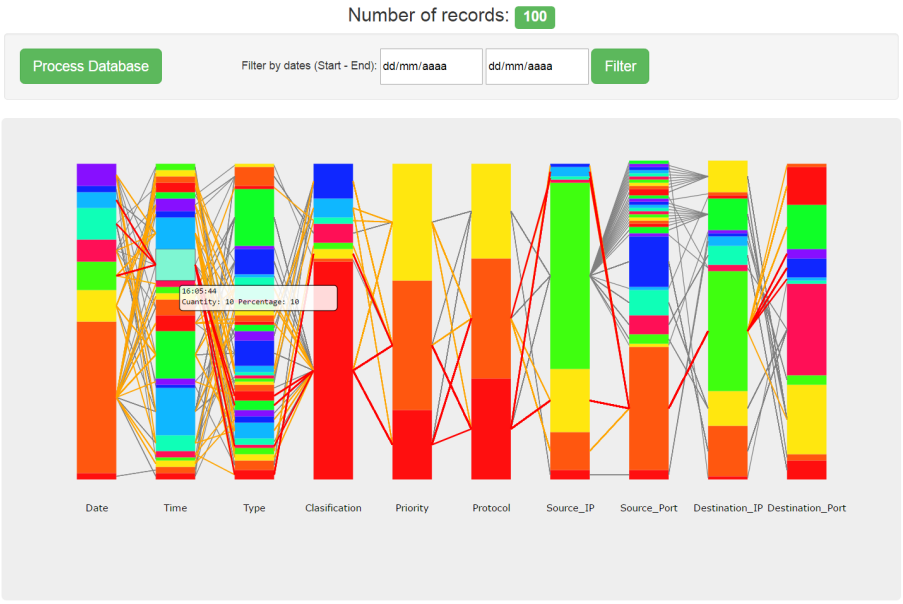

Figura 2. Selección de un vértice

En la Figura 3 se ve el funcionamiento del filtrado por fechas. Es capaz de seleccionar un rango de fechas y mostrar en pantalla los diferentes valores que se presentan en ese rango. Se aprecia en la imagen que se seleccionó uno de los puertos de origen y se resaltó la relación fuerte que hay entre ese valor y una hora específica, sin dejar de lado los demás caminos posibles que existen. Se observa cómo la herramienta permite las mismas funcionalidades con los registros filtrados que si el análisis se hace de todos los registros.

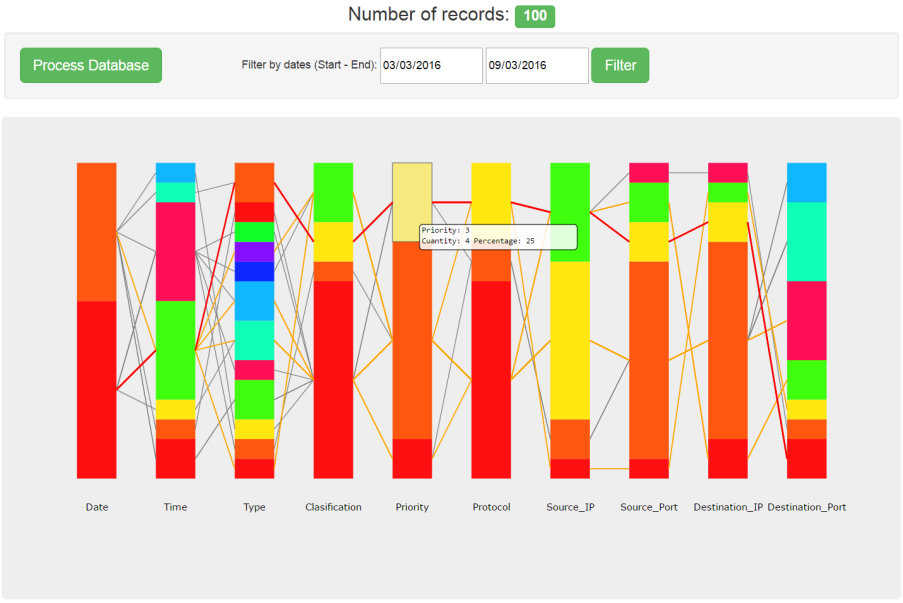

Figura 3. Filtrado por fechas

4. Línea Futura y recomendaciones

El desarrollo del sistema no es del todo automático, ya que la información capturada debe poseer un formato específico y ser almacenada en una sola dirección. La base datos almacena una vasta cantidad de registros sobre las trazas pero el aplicativo no puede manipularlos en el momento de presentar los gráficos.

Por lo tanto, la siguiente lista señala los puntos de la trayectoria que se podría seguir:

- Obtener nuevos datos desde diferentes formatos de archivo.
- Insertar registros en la base de datos utilizando más de un archivo a la vez.
- Usar un alto volumen de datos.
- Adición de mapas geográficos.

El resultado de este proyecto es el primer paso para sentar las bases de la herramienta y poder agregar en el futuro más posibilidades para realizar filtros, tomando como ejemplo el filtrado por fechas que se ha aplicado en esta oportunidad. Como alternativa para el desarrollo futuro puede pensarse en filtros por direcciones IP, puertos y horas, entre otros.

Como recomendación para el seguimiento y eventuales mejoras en el diseño de esta herramienta, se propone no utilizar otros frameworks debido a que dificulta la integración y desarrollo.

5. Conclusiones

Con lo expuesto se ha demostrado que el uso de gráficos, ya sean de barras, circulares o mapas geográficos, colabora a que el lector tenga un mayor grado de entendimiento de la información, especialmente cuando esta acumula altos volúmenes de datos. Esto hace que dichas representaciones jueguen un papel preponderante en la toma de decisiones y hacen innecesario el uso de textos amplios de difícil manejo y comprensión.

Durante la elaboración del proyecto se enfrentaron diversos retos que motivaron a utilizar técnicas como minería de datos y uso de algoritmos matemáticos. Estas técnicas ayudaron a superar los retos, así como a consolidar el conocimiento del equipo de trabajo en estas áreas.

A continuación, se presentan los puntos específicos en los que se concluye con el desarrollo de este proyecto:

- Se confirma que la visualización de la información permite a los Analistas de Seguridad detectar vulnerabilidades más rápidamente y tomar decisiones en un tiempo más corto.
- Se reducen los tiempos de respuesta a partir del análisis que se genera en tiempo real.
- La visualización de la información ayuda notablemente a la comprensión de grandes masas de datos.
- Se pueden detectar patrones y comportamientos de ataques; por ejemplo, tipos de ataques que se realizan con más frecuencia y la fecha y hora en que se generan.
- Se determina que al tener mayores herramientas de visualización se puede simplificar el trabajo de los Analistas e Ingenieros de Seguridad en vista de lo amigable que resulta revisar un gráfico, en contraposición con la complejidad y dificultad que implica el análisis basado en un documento de texto plano de datos.
- Por último, a partir de una única fuente de datos se pueden aplicar diferentes tipos de visualización y filtros, lo que permite realizar diferentes tipos de análisis y con mayores niveles de profundidad.

Referencias

Beom-Hwan, y Yoon, C. (2011). An efficient network attack visualization using security quad and cube. *ETRI Journal*, *33*(5), 770 - 779. pages 5

Bhardwaj, A., y Singh, M. (2015). Data mining-based integrated network traffic visualization framework for threat detection. *Neural Computing And Applications*, *26*(1), 117-130. pages 3

Chavhan, s., Sushil kumar1, y Nirkhi, s., S. M.2. (2012). Visualization techniques for digital forensics: A survey. *International Journal of Advanced Computer Research*, *2*(6), 74 - 78. pages 4

Ennert, M., Mados, B., y Dudláková. (2014). Data visualization of network security systems. *Acta Electrotechnica e Informatica*, *14*(4), 13 - 16. pages 5

Fry, B. (2016). *The data visualization process.* urlhttps://www.dashingd3js.com/the-data-visualization-process. pages 3

Janowski, P., Lucjan1and Owezarski. (2010). Assessing the accuracy of using aggregated traffic traces in network engineering. *Telecommunication Systems*, *43*(3/4), 223-236. pages 4

Luse, A. (2008). Framework for visualization of intrusion detection events. *Information Security Journal: A Global Perspective*, *17*(2), 95 - 107. pages 5

Manovich, L. (2010). What is visualization? *paj: The Journal of the Initiative for Digital Humanities, Media, and Culture*, *2*(1). pages 3

Roesch, M., y cols. (1999). Snort: Lightweight intrusion detection for networks. En *Lisa* (Vol. 99, p. 229-238). pages 4

Verbert, k., Katrien, Govaerts, E., Stenand Duval, Santos, Assche, Parra, y Klerkx. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, *18*(6), 1499 - 1514. pages 4