Representación visual de patrones de ataque en ciberseguridad

Edgardo Antonio Rojas Arias y Dennis Coto Leiva

Escuela de Ingeniería,
Universidad Latinoamericana de Ciencia y Tecnología,
ULACIT, Urbanización Tournón, 10235-1000
San José, Costa Rica
erojasa612,dcotol848@ulacit.ac.cr
http://www.ulacit.ac.cr

Resumen Debido a la enorme cantidad de ataques a las redes de diversos tipos de organizaciones se requiere determinar los diferentes patrones y ataques que se realizan. Con base en esto, ha surgido la necesidad de procesar, analizar y visualizar los datos que se derivan de los procesos de ciberseguridad, para permitir su interpretación y la toma de decisiones para mejorar los niveles de protección de la información. Como consecuencia, este trabajo de investigación propone el diseño de una herramienta de visualización para facilitar el análisis de información de ciberseguridad de forma interactiva y dinámica.

Keywords: Visualización de seguridad, ciberseguridad, Big Data

1. Introducción

Con el aumento de las tecnologías de información, se ha incrementado la necesidad de llevar a cabo el procesamiento y análisis de datos para facilitar la toma de decisiones usando información producida en tiempo real. Esto implica obtener los datos, realizar el análisis de estos y utilizar visualización para proporcionar resultados, de forma intuitiva, simple e interactiva, a los usuarios.

En términos generales, Big Data implica la recolección de grandes volúmenes de datos, su análisis y correlación, así como la identificación de distintos patrones lógicos que permiten convertirlos en conocimiento (Boyd y Crawford, 2012). El desarrollo de herramientas de visualización permiten el análisis de forma dinámica e interactiva para facilitar que los usuarios trabajen de forma más precisa en el análisis de datos.

Como consecuencia, el objetivo de este trabajo de investigación es apoyar el análisis de grandes volúmenes de datos mediante la representación visual de relaciones entre elementos y posibles patrones de ataque en el campo de la ciberseguridad. Por lo que en las siguientes secciones se lleva a cabo una revisión una revisión bibliográfica sobre el uso de la visualización de información en el análisis de datos, se discute el diseño de una propuesta de visualización para el análisis de relaciones entre elementos en contextos diversos, se presenta un caso de estudio y se realizan las conclusiones de la investigación.

2. Antecedentes

Las representaciones visuales han ido evolucionando con el tiempo, debido a su facilidad de representar información precisa y de fácil entendimiento para las demás personas, esto debido a la necesidad de realizar análisis de grandes volúmenes de dat. Gracias a esta necesidad las herramientas visuales han alcanzado un punto de evolución constante sobre el como se desea observar y manipular dicha información.(Card y Mackinlay, 1997)

Las herramientas visuales sirven como medios para expresar o comunicar a través de diferentes visualizaciones la información que se proceso y analizó con anterioridad, esto con el fin de tener un mejor entendimiento de la información que permita la mejora de procesos, la toma de decisiones o de acciones a tomar con respecto a la información analizada.

Existen diferentes tipos de herramientas visuales y como se se mencionó anteriormente, estas han ido evolucionando de acuerdo a las necesidades y a la gran cantidad de datos que se tienen que procesar actualmente.

Con el surgimiento del Big Data(Jacobs, 2009), que hace referencia a los grandes volúmenes de datos que se almacenan y que cuentan con infinidad de patrones que permiten relacionar la información para así realizar análisis detallados de dichos datos, las herramientas visuales han logrado manipular y procesar toda esta información de manera sencilla y dinámica permitiendo así utilizar de manera más precisa y exacta dicha información.

3. Diseño de la herramienta

El desarrollo de esta aplicación, requirió de varias etapas, con las cuales se analizaba y se adaptaba la herramienta de acuerdo a las necesidades que se requerían.

3.1. Adquisición de datos para la carga de base de datos

En esta etapa se procedió a estudiar varios conjuntos de datos, tomando como referencia el número de columnas presentes, las filas de información y la presencia o ausencia de errores. Con ese propósito se seleccionaron los datos más relevantes y comprensibles.

Los datos seleccionados es un conjunto de datos de SNORT que contiene las direcciones IP y puertos atacados en Internet, de un caso sucedido años atrás. Este conjunto de datos se encuentra disponible para su estudio o análisis de forma libre, aunque cierta información fue censurada o eliminada por razones de confidencialidad.

3.2. Extracción y limpieza de los datos

Esta etapa del proceso contempló los siguientes pasos:

1. Limpieza de errores: los registros que no cumplían con las especificaciones necesarias, principalmente que no estuvieran completas fueron eliminadas. Lo anterior tomando en cuenta la ausencia de valores debido a la supresión que fue realizada por motivos de confidencialidad.
2. Conversión de formato: el conjunto de datos fue convertido a formato JSON para poder cargarlo en MongoDB.

3.3. Carga de datos

La carga de los datos requirió que el conjunto fuera dividido en grupos de aproximadamente 500 registros por bloque. Lo anterior debido a que durante el proceso de conversión la herramienta utilizada no pueden procesar tantos datos a la vez. Por lo que se crearon una serie de bloques de datos que luego fueron en MongoDB.

3.4. Modelado de sistema

El desarrollo de la herramienta se basó en el patrón Modelo-Vista-Controlador (MVC), el cual utiliza varias capas para separar el modelo de la datos de la lógica del negocio y la interfase del usuario. Como consecuencia, la herramienta está estructura con una capa de modelo de datos, que permite la conexión y consulta a la base de datos, la capa de comunicación (denominada como controlador) y la capa de visualización, donde se encuentran las representaciones visuales.

3.5. Detalles del Diseño de la visualización

La implementación de la visualización se realizó utilizando Mongodb (base de datos) la cual es una base de datos NoSQL , lo que significa que no realiza las típicas consultas SQL que se pueden ejecutar en bases de datos como Oracle, SQL Server o MySQL, esta es la gran diferencia con respecto a las demás bases de datos relacionales que se usan diariamente en la mayoría de las empresas.

Las bases de datos relacionales funcionan mediante registros los cuales se pueden entender como líneas de datos los cuales deben de cumplir un patrón determinado para poder ser guardados, si este patrón no se cumple es imposible registrar los datos, sin embargo la versatilidad de las bases de datos NoSQL en este caso MongoDB registran los datos en forma en documentos, a que nos referimos con esto, pues bien un documento puede contener una colección de datos semejantes sin embargo no es necesario un patrón patrón de elementos.

Además se utiliza Express, esta es una infraestructura que permite un desarrollo rápido de aplicaciones web debido a que proporciona herramientas que permiten el rápido desarrollo de aplicaciones, lo cual es ideal para el proyecto debido al poco tiempo que se tiene para implementar la aplicación de visualización, además Express funciona de manera ágil con node.js la cual es un intérprete de JavaScript (lenguaje de programación que funciona de lado cliente para el desarrollo de aplicaciones)

Finalmente la herramienta que nos ayudó en el proceso propio de la visualización es Highcharts, esta es una librería de objetos hechos en JavaScript para la realización de visualizaciones, esta librería se adaptó a nuestro proyecto, se modificó su código para ser adaptada a la aplicación, además de ello se realizó un código extra que ayuda a la aplicación en el despliegue de la información (ver figura 3).

4. Caso de Estudio

El conjunto de datos utilizado fue recolectado haciendo uso de Snort Traffic (Roesch y cols., 1999), el cual es un sistema de análisis de paquetes, que detecta y registra intrusiones y ataques de red, tales como desbordamientos de buffer y ataques CGI, entre otros.

La herramienta es una aplicación de una sola página[1] que cuenta con dos tipos de visualización: la primera visualización es un contador de datos (ver Figura 1), mientras que la segunda visualización permite relacionar los datos de forma lineal (ver Figura 2).

La visualización de contadores permite observar la cantidad de repeticiones de un valor en todo el conjunto de datos o bien la cantidad de repeticiones de una búsqueda particular, en tanto que la visualización para relacionar los datos de forma lineal ofrece la posibilidad de analizar diferentes patrones y relaciones que se obtienen al efectuar una búsqueda, lo cual permite llevar a cabo el análisis de los patrones de ataque.

De forma adicional, al seleccionar algún valor de cualquiera de las visualizaciones, se desplegará al lado derecho de la pantalla un cuadro negro, el cual cuenta con la descripción de los datos. En el caso de la Figura 1, se muestra la cantidad de repeticiones que existen sobre un valor consultado y en la Figura 2, se puede ver la información relacionada con el punto señalado (ver Figura 3).

[1] Single Page Application o SPA por su acepción en inglés.

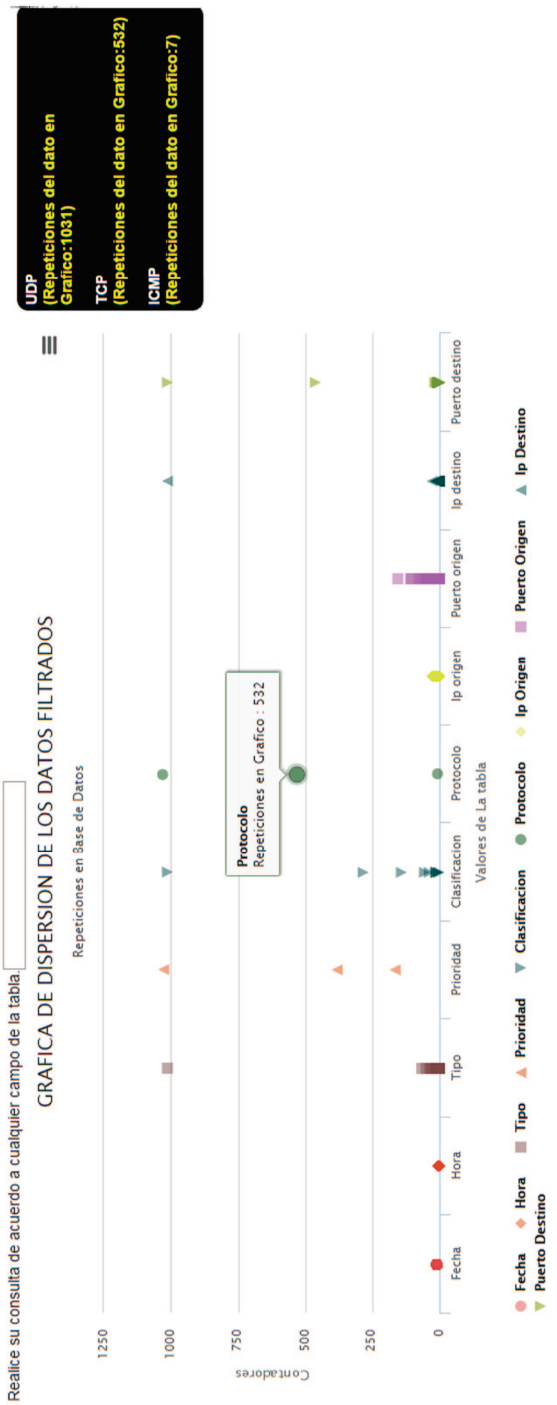

Figura 1. Vista principal de la visualización de contadores.

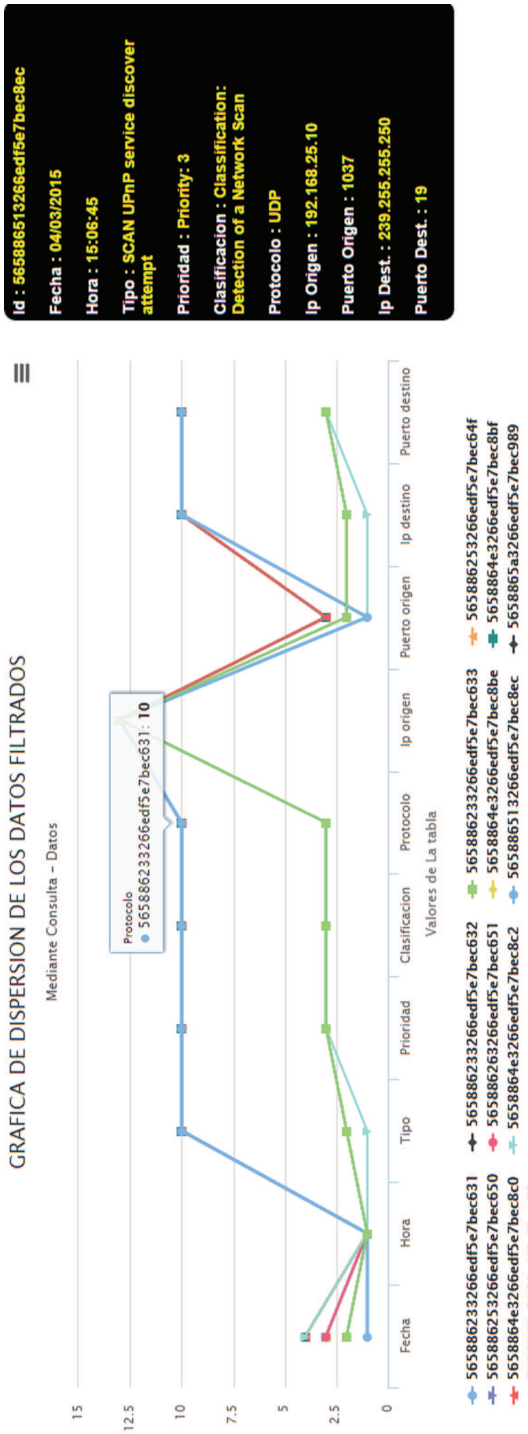

Figura 2. Relación de posibles patrones de ataque.

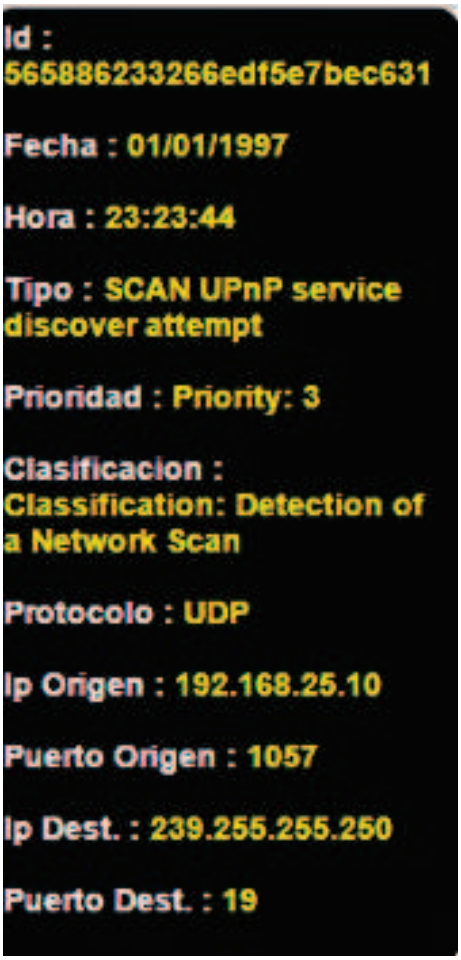

Figura 3. Detalle de un elemento.

5. Conclusiones

Los resultados obtenidos durante el desarrollo del proyecto, muestran la gran cantidad de ataques que se realizan diariamente a través de la red, y de cuan importante es poder llevar a cabo análisis de toda esta data, con la aplicación que se desarrolló se puede observar de manera muy dinámica el como utilizar estos datos, para así procesarlos, analizarlos y crear reportes a través de las diferentes visualizaciones desarrolladas.

Con dichos reportes se podrá observar de forma más práctica los distintos patrones de ataque, así como el origen de estos, al visualizar esta información, se podrá bloquear los orígenes de ataque, mejorar la seguridad de los puertos más atacados o bien al observar por medio de estas visualizaciones cuales son las debilidades de la red, tomar mejores decisiones en la seguridad de estas y de los sistemas con el fin de prevenir la fuga de información y daños provocados a los equipos.

Adicionalmente dichas visualizaciones son fáciles de adaptar a los diferentes tipos de datos que maneje cada entidad por lo que puede ser de mucha utilidad para diferentes organizaciones que requieran análisis precisos y dinámicos sobre la seguridad de sus redes.

Las herramientas visuales son muy utilizadas diariamente para la observación de ataques que se dan a través de la red, como lo es Norse Attack Map que muestra los ataques en tiempo real que se dan entre distintos países, de ahí la importancia de utilizar herramientas visuales para el análisis de estos ataques, esto con el fin de prevenir y mejorar nuestros sistemas de seguridad.

Referencias

Boyd, D., y Crawford, J. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662-679. Descargado de http://dx.doi.org/10.1080/1369118X.2012.678878 doi: 10.1080/1369118X.2012.678878 pages 1

Card, S., y Mackinlay, J. (1997, Oct). The structure of the information visualization design space. En *Information visualization, 1997. proceedings., ieee symposium on* (p. 92-99). doi: 10.1109/INFVIS.1997.636792 pages 2

Jacobs, A. (2009, agosto). The pathologies of big data. *Commun. ACM*, *52*(8), 36–44. Descargado de http://doi.acm.org/10.1145/1536616.1536632 doi: 10.1145/1536616.1536632 pages 2

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