Identificación visual de ciberataques mediante técnicas de reducción de la dimensionalidad y visualización de la información

Nicolas R. Enciso[1], Jorge E. Camargo[2]

Universidad Nacional de Colombia sede Bogotá, Grupo de investigación UNSecure Lab.

nricardoe@unal.edu.co[1], jecamargom@unal.edu.co[2]

Resumen. Debido a los grandes avances de Internet y la computación, la sociedad está cada vez más conectada digitalmente mediante diferentes tipos de dispositivos y sistemas de información. Esto hace que se generen mayores riesgos de recibir ciberataques. En este trabajo se presenta un método para identificar automáticamente ciberataques a partir del contenido de los headers de paquetes HTTP. El método utiliza un conjunto de datos de 60.000 ítems en el que se cuenta con paquetes normales y con paquetes de diferentes tipos de ciberataques tales como XSS, SQL Injection y CRLF. Los resultados muestran que es posible visualizar el conjunto de datos en un espacio de 2 dimensiones para resaltar visualmente cuáles ítems son ciberataques.

Keywords: PCA, URL, CRLF, XSS, SQL injection, auto valor, auto vector.

1. Introducción

En la creciente sociedad conectada, el número de ataques informáticos ha ido en exponencial aumento, poniendo en el centro de la discusión social la seguridad de la información [1][2]. Con la cantidad de nuevas amenazas que se van descubriendo, es necesario construir nuevas herramientas que permitan detectar automáticamente ataques al mismo ritmo en el que aparecen. En ese contexto es donde los algoritmos de machine learning entran a jugar un papel importante debido a su gran capacidad para extraer patrones a partir de los datos, como en el caso de los sistemas de detección de intrusiones (IDS) [3] [4], con los cuales se pueden obtener tasas de detección de casos anómalos con porcentajes considerables de precisión, además de usar técnicas de big data combinadas [5]. Sin embargo, para poder obtener tales porcentajes de precisión, es indispensable hacer un tratamiento previo de los datos que serán usados como entrada al algoritmo de machine learning.

La extracción de características es un paso fundamental en los procesos de machine learning, del cual depende el éxito de la aplicación de los mismo. Junto con la correcta selección de los campos del dataset a usar, se debe hacer un análisis de qué características hacen diferenciador los grupos en los que se quieren clasificar los casos del dataset. Adicionalmente, cuando se obtiene una gran cantidad de variables extraídas como características en cada caso, se hace complejo su análisis, debido a la

imposibilidad de tener una imagen gráfica que represente de manera fiel el comportamiento de los datos, así como la comprobación de haber realizado una elección de características correcta.
En la extracción de características a partir de URLs, se pueden detectar casos de ataques tales como CRLF, XSS, SQL injection e incluso phishing **[6][7]**, capturando sus principales características con un análisis de las diferentes partes que componen una URL. En cuanto al problema de la imposibilidad de lograr visualizar y simplificar los datos luego de la extracción de características, existen varias técnicas capaces de poder reducir el número de dimensiones de los datos, sin perder información de manera importante, como es el caso de la factorización no negativa de matrices (NMF) el cual ha sido usado con éxito en la detección de *outliers* en sensores **[8]** y la detección de anomalías en tarjetas inteligentes **[9]**, y PCA, el cual ha sido usado con éxito en la clasificación de casos de anomalías en sistemas de detección de intrusiones **[10].**

2. Materiales y método

Para la metodología desarrollada, se usan varios componentes de software. Python como lenguaje de programación básico, el cual hace uso de las librerías *Scipy* para el tratamiento de datos, *Pandas* para manejo de tablas, *Numpy* para tratamiento matemático, *Matplotlib* para la visualización y *Scikit-learn* para herramientas de probabilidad, normalización de datos, así como la herramienta encargada de extraer autovalores y autovectores de la matriz de datos.

En la metodología desarrollada se busca tener respuestas a las siguientes preguntas:

1. ¿Es posible a partir de URLs en *headers* de http, identificar ataques?
2. ¿Aplicando reducción dimensional de variables en un dataset, se puede obtener una diferenciación o agrupación de tipos gráficamente que demuestre tener una correcta caracterización?

La primera pregunta de investigación es abordada desde el punto de vista de la extracción de características. Teniendo URLs provenientes de casos de ataque y neutrales, se plantea la posibilidad de identificar ataques por medio de análisis de las URLs. En cuanto a la segunda pregunta, se quiere ver si es posible tener una clara imagen de los dos tipos de datos que se presentan en el dataset, luego de la extracción de características que permitan identificar *clusters* de un mismo tipo.

2.1. Extracción de características

El dataset usado proviene de un proyecto realizado por el gobierno de España, a través del Consejo Nacional de Investigación, en el Instituto de seguridad de la Información **[11]**, el cual contiene más de 60.000 HTTP headers o web requests, en los cuales hay 25.000 casos de ataques, donde se presentan ataques CRLF, SQL injection, y XSS. El dataset ha sido usado con éxito en otros trabajos para la detección de anomalías en ataques web **[12]**. Se toma del dataset la URL de cada web request.

La extracción de características se hace basado en la caracterización de URLs para detección de phishing aplicado por un trabajo previo **[7]**, y en la identificación de características propias de los tipos de ataque que contiene el dataset, y que son identificables en las URLs.
Las 8 características extraídas son las siguientes: longitud, cantidad de caracteres especiales (@,-,#,$), cantidad de palabras sospechosas (confirm, account, admin, etc), cantidad de palabras del lenguaje SQL (SELECT, WHERE, etc), cantidad de palabras referidas a XSS (alert, param script, etc.), cantidad de caracteres propios de ataques CRLF (%,0), y dos métricas de distribuciones de probabilidad (prueba no paramétrica de Kolmogorov-Smirnov, medida no simétrica de divergencia de Kullback-Leibler) entre la distribución de letras en el idioma español y en el dado por la URL a estudiar. De esta manera se extraen características que permiten identificar si la URL tiene rasgos de un ciberataque. La extracción de características se aplica a los datos de ambos tipos (ataque y neutral), obteniendo un nuevo dataset en formato CSV con valores numéricos, y una etiqueta final definiendo el tipo al que pertenece. Ambos tipos de dato son puestos en un único archivo.

2.2. Reducción de la dimensionalidad

Teniendo el dataset caracterizado, se procede a la reducción de dimensionalidad con PCA. Para ello, se toma el dataset y se parte en dos usando la librería *Pandas*, una con los valores numéricos, los cuales componen 8 dimensiones de variables, y la otra con sólo las etiquetas del tipo al que pertenece el caso. Posteriormente, se aplica normalización en la matriz de características, donde se elimina la media y se escala a una varianza unitaria. Se continúa calculando la matriz de covarianza usando *Numpy*, obteniendo así una matriz cuadrada *nxn* con *n* la cantidad de dimensiones o características, en este caso 8. Con la matriz obtenida, se calcula la matriz de autovectores con dimensión nxn, y la de autovalores de *1xn*. Cada columna de la matriz de autovectores está relacionada en el mismo orden, con cada columna de la matriz de autovalores, de manera que se procede a organizar, manteniendo dicho orden de par columna autovector-autovalor, de mayor a menor valor según el autovalor. Se hace ese ordenamiento debido a que los autovalores muestran la cantidad de varianza explicada que contiene la variable, por lo cual, explica de mayor forma los cambios que se encuentran entre los casos del dataset, además de la propia característica de un autovector asociado, de mantener una misma dirección, luego de aplicar varias transformaciones lineales, como lo son la extracción de la matriz de covarianza del dataset original.
Finalmente, se eligen los 2 primeros autovalor-autovector, para poder visualizar el comportamiento de los datos en 2 dimensiones; cada dimensión es un PCA. La varianza que contienen las 2 elegidas, como se puede ver en la Figura 1

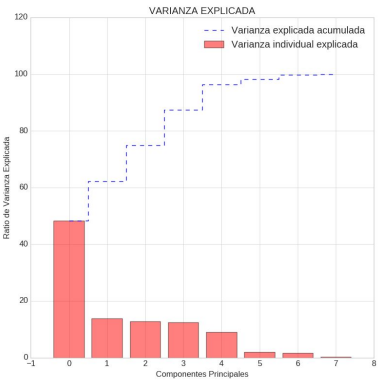

Figura 1. Porcentaje acumulado de varianza de cada componente principal.

2.3. Visualización en 2D

Con los dos componentes principales elegidos, se multiplica la matriz 8x2 de los autovectores correspondientes a los dos autovalores mayores, con la matriz original estandarizada de dimensiones *60.000x8*, en la que 60.000 corresponde al total de casos del dataset. Una vez se extraen los 2 componentes principales, se obtiene una matriz de *60.000x2*, la cual es la matriz final con los puntos a graficar. De esta matriz resultante se utiliza el componente 1 como *x* y el componente 2 como *y*, de tal forma que cada ítem del conjunto de datos estará representado por las coordenadas *(x,y)*, con lo cual se busca responder pregunta de investigación 2. En la Figura 2 se ve el resultado final (azul para ataques, rojo para neutral).

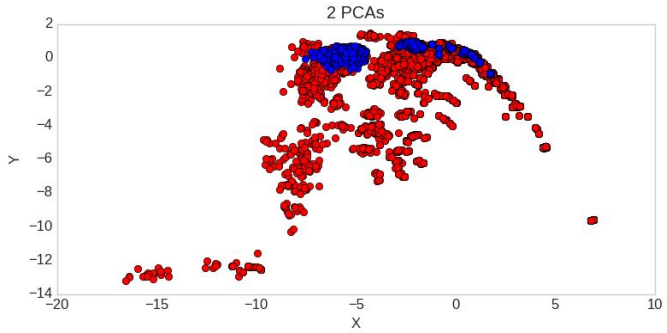

Figura 2. Visualización de los datos utilizando los 2 componentes principales.

3. Conclusiones

El método permite identificar visualmente cuáles paquetes son normales y cuáles son ciberataques. Con la simplificación del dataset, y una pérdida no significativa en información, se puede continuar con la etapa de uso de los datos preparados en el presente trabajo, como entrada para algoritmos de clasificación en machine learning. Como trabajo futuro se plantea utilizar métodos de Machine Learning tales como clasificación y agrupamiento.

Referencias

1. Verizon 2017 Data Breach Investigations Report (DBIR,2017),Revisado en October 13, 2018 de http://www.verizonenterprise.com/verizon-insights-lab/dbir/2017/
2. Idhammad, M., Afdel, K. & Belouch, M. Semi-supervised machine learning approach for DDoS detection. Applyied Intelligence, (2018) 48: 3193.

3. Sanjay Kumar, Ari Viinikainen, Timo Hamalainen, «Machine learning classification model for network based intrusion detection system», in 11th International conference for internet technology and secured transactions (ICITST), 2016.
4. Constantinos Kolias, Georgios Kambourakis, Angelos Stavrou, «Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset », IEEE Encuestas de comunicaciones y tutoriales volúmen 18, Universidad del Egeo, Samos, Grecia, 2015.
5. Kalyan Veeramachaneni, Alfredo Cuesta-Infante, Vamsi Korrapati, Costas Bassias, Ke Li Ignacio Arnaldo. (2016). AI 2: Training a big data machine to defend de MIT Massachusetts Institute of Technology.
6. Doyen Sahoo, Chenghao Liu, and Steven C.H. Hoi. (2017). Malicious URL Detection using Machine Learning: A Survey de Universidad de Administración de Singapur, Escuela de sistemas de información.
7. Alejandro Correa Bahnsen, Eduardo Contreras Bohorquez, Sergio Villegas, Javier Vargas, Fabio A. González. (2017). Classifying Phishing URLs Using Recurrent Neural Networks. de Easy Solutions Research, MindLab Research Group Universidad Nacional de Colombia.
8. Hamoud Alshammari, Oussama Ghorbel, Mohammed Aseeri and Mohamed Abid. (2018). Non-Negative Matrix Factorization (NMF) for outlier detection in Wireless Sensor Networks de Universidad Al Jouf (Reino de Arabia Saudí), CES centro de investigación Universidad de Sfax (Túnez).
9. Emeric Tonnelier, Nicolas Baskiotis, Vincent Guigue, Patrick Gallinari. (2014). Anomaly detection and characterization in smart card logs using NMF and Tweets de Universidad de París, La Sorbonne, París Francia.
10. Mei-Ling Shyu, Shu-Ching Chen, Kanoksri Sarinnapakorn, LiWu Chang. (2003). A Novel Anomaly Detection Scheme Based on Principal Component Classifier de Universidad de Miami, Laboratorio de investigación Naval en sistemas computacionales Washington DC.
11. Carmen Torrano Giménez, Alejandro Pérez Villegas, Gonzalo Álvarez Marañón. (2010). HTTP DATASET CSIC 2010. Octubre 12 2018, de Consejo Nacional de Investigación, Instituto de seguridad de la Información, gobierno de España Sitio web: http://www.isi.csic.es/dataset/
12. C. Torrano-Gimenez, A. Perez-Villegas, G. Alvarez. An anomaly-based approach for intrusion detection in web traffic. Journal of Information Assurance and Security, vol. 5, issue 4, pp. 446-454. ISSN 1554-1010 (2010).