EXTRACCIÓN DE CONOCIMIENTO A PARTIR DE FUENTES DE DATOS REALES PROCEDENTES DE LA MONITORIZACIÓN DE EVENTOS DE SEGURIDAD

Bravo Gómez, Alberto

La Gestión de Eventos e Información de Seguridad (Security Information and Event Management - SIEM) está cada vez más implantada en las organizaciones, debido a la importancia que en los últimos años ha adquirido la seguridad en los Sistemas Informáticos. Mediante estos sistemas de seguridad las organizaciones pueden detectar amenazas, vulnerabilidades y estimar riesgos de seguridad. La gestión de eventos e información relacionada con la seguridad se realiza mediante sistemas comerciales que facilitan toda la información, procesando diferentes fuentes de datos. Este capítulo presenta un novedoso sistema de extracción de conocimiento basado en la monitorización de eventos de seguridad que permite complementar la información de los sistemas comerciales y también predecir conductas futuras de riesgo.

1. Introducción

En este artículo se presenta un completo sistema de extracción de conocimiento a partir de datos reales anonimizados, que persigue una doble finalidad con el desarrollo de este trabajo: (i) completar y complementar la monitorización proporcionada por el SIEM de la organización; y (ii) predecir conductas futuras que permitan la anticipación a potenciales situaciones de riesgo, con una precisión bastante fiable.

Se consideran tres fuentes de datos diferentes que constan de 10.000 tuplas para esta primera fase:

- Microsoft: Archivo con información sobre servidores de Windows que están monitorizados.
- Firewalls: Archivo con información sobre el cortafuegos o firewall utilizado (Fortinet, 2018).
- Antivirus: Archivo con información sobre los eventos capturados por el antivirus utilizado, concretamente Symantec Endpoint Protection de la empresa Symantec.

Todos los archivos se encuentran en formato CSV y con una subestructura interna denominada SYSLOG (Internet Engineering Task Force, 2009) asignada por el SIEM QRadar utilizado por Viewnext.

2. Sistema de Extracción de Conocimiento

El sistema desarrollado consta de 4 fases de implementación bien diferenciadas, donde finalmente se determina el impacto y severidad para cada uno de los eventos analizados.

2.1. Fase 1: Preprocesado y Formateo de Datos

El preprocesado y formateo de datos procedentes de estos tres archivos se fundamenta en un proceso semiautomático, ya que requiere de cierta intervención manual. Principalmente se basa en la eliminación de información prescindible, así como solventar irregularidades que algunos archivos poseen. También se han realizado sintetizaciones en la información que presentan, para facilitar las fases posteriores de análisis.

2.2. Fase 2: Categorización y Clasificación de Amenazas

Se realiza una búsqueda sobre diferentes patrones, métodos y estándares que permiten la detección de amenazas (Jouini, Rabai y Aissa, 2014). Finalmente, la clasificación se realiza siguiendo el modelo STRIDE (Shostack, 2014), que categoriza el impacto de las amenazas en seis grupos: ***S***uplantación de identidad, ***T***ampering o manipulación de información, ***R***epudio, divulgación de ***I***nformación, ***D***oS y ***E***scalada de privilegios. Se distinguen a su vez otras tres subfases o subprocesos para la categorización y clasificación de amenazas.

2.2.1. Identificación de las Amenazas

El sistema implementado identifica amenazas en los archivos considerados, analizando campos concretos para identificarlas y clasificarlas. Algunos de esos campos se refieren al identificador de evento, *log* o *sid*, permitiendo fácilmente diferenciar entre falsos positivos y falsos negativos.

2.2.2. Categorización de las Amenazas

El sistema propuesto realiza la categorización de las amenazas según el proceso descrito por Adam Shostack (Shostack, 2014). Como resultado de este proceso de categorización, a cada evento identificado se le asigna una de las categorías que se cumplen en STRIDE, como se muestra de ejemplo en la Tabla 1.

Tabla 1. Ejemplo del resultado de la categorización de amenazas.

Amenaza Identificada	S	T	R	I	D	E
SMB Double Pulsar Ping	X	X		X		X
Fake TechSupport Domains	X	X		X		
Microsoft Windows SMB Remote Code	X	X	X	X		X
Microsoft SMB Disclosure Attempt	X	X	X	X	X	X

Fuente: Elaboración propia a partir de Bravo, Sancho y Caro (2019).

2.2.3. Asignación de Criticidad

Tras la categorización de amenazas, el modelo propuesto establece un sistema de puntuación para determinar el grado de criticidad. De este modo, el sistema se encarga de asignar a cada amenaza su impacto, en función a los valores establecidos en la Tabla 2. Puede apreciarse que el valor de criticidad asignado (Alto, Medio o Bajo) se establece en función del impacto determinado en STRIDE.

Tabla 2. Impacto de las amenazas con su criticidad designada.

IMPACTO STRIDE	CRITICIDAD
1 – 2	Bajo
3 – 4	Medio
5 – 6	Alto

Fuente: Elaboración propia a partir de Bravo *et al.* (2019).

2.3. Fase 3: Análisis de Datos

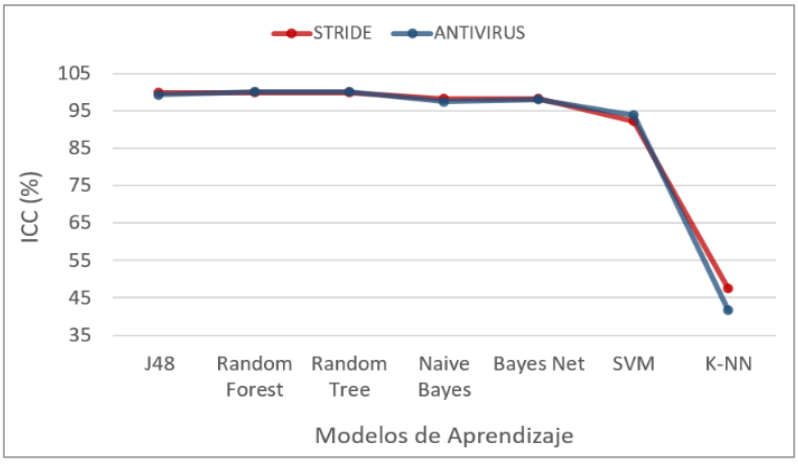
El análisis de datos se realiza mediante la herramienta Weka (Witten y Frank, 2011) para completar la etapa de extracción de conocimiento. Este entorno permite la ejecución de algoritmos y modelos de aprendizaje de forma sencilla y flexible. El objetivo se basa en poder establecer una categorización y clasificación automática de amenazas mediante modelos de aprendizaje. Se han realizado pruebas experimentales en base a los clasificadores más habituales: J48/C4.5, Random Forest, Random Tree, Naive Bayes, Bayes Net, LibSVM y SimpleKMeans. Cross-validation establece cómo serán ejecutados los clasificadores en la herramienta Weka.

2.4. Fase 4: Análisis de Resultados

Esta fase se centra en estudiar los resultados obtenidos sobre los 3 archivos analizados (Antivirus, Firewalls y Microsoft). Debido a la limitada extensión de este artículo, se mostrarán únicamente los resultados más

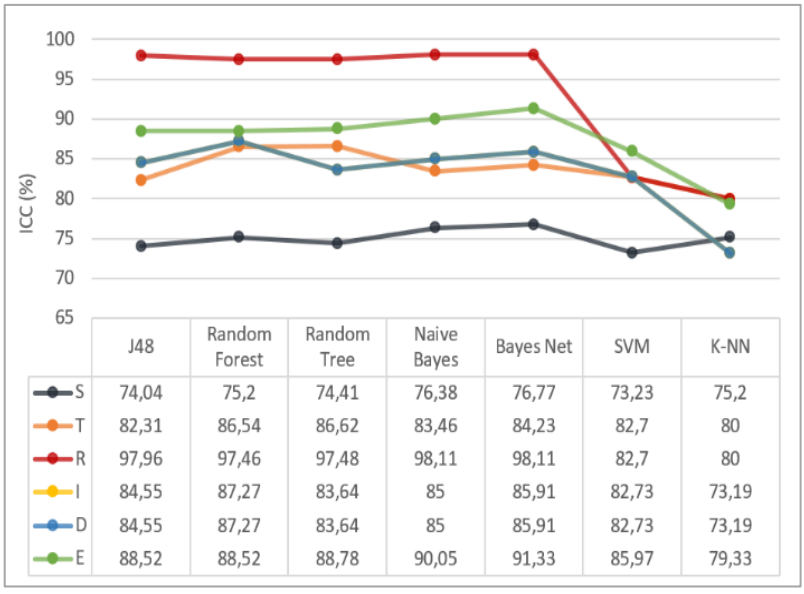
relevantes obtenidos en el archivo Antivirus. No obstante, se desea hacer constar que los resultados conseguidos son muy similares en los otros 2 archivos analizados (Firewalls y Microsoft).

En la Gráfica 1 se visualizan los resultados obtenidos con los modelos de clasificación en base a la severidad asignada con STRIDE y a la aportada por el antivirus comercial. Se puede apreciar una gran similitud en los valores de ICC (Instancias Correctamente Clasificadas), siendo levemente mejores en los resultados obtenidos por la severidad de STRIDE, especialmente con los modelos Naive Bayes y K-NN.



Gráfica 1. Efectividad entre la criticidad propuesta (STRIDE) y la del proveedor de Antivirus. Fuente: Elaboración propia a partir de Bravo *et al*. (2019).

Otro aspecto importante del desarrollo del trabajo se basaba en determinar las diferentes categorías de STRIDE que se aplican en una amenaza concreta, como muestran los resultados de la Gráfica 2.



	J48	Random Forest	Random Tree	Naive Bayes	Bayes Net	SVM	K-NN
S	74,04	75,2	74,41	76,38	76,77	73,23	75,2
T	82,31	86,54	86,62	83,46	84,23	82,7	80
R	97,96	97,46	97,48	98,11	98,11	82,7	80
I	84,55	87,27	83,64	85	85,91	82,73	73,19
D	84,55	87,27	83,64	85	85,91	82,73	73,19
E	88,52	88,52	88,78	90,05	91,33	85,97	79,33

Gráfica 2. Efectividad entre las categorías de STRIDE balanceadas de Antivirus. Fuente: Elaboración propia a partir de Bravo *et al.* (2019).

Las clases de las categorías se encontraban originalmente muy desbalanceadas, por lo que se realizó un balance para poseer las mismas muestras de casos positivos y negativos. De esta manera, los clasificadores no muestran predominancia en los resultados a favor de la clase mayoritaria. Los resultados obtenidos por los clasificadores son generalmente buenos, destacando aquellos pertenecientes a la categoría R.

Mediante los resultados obtenidos se pone de manifiesto el buen desempeño que posee esta primera versión del sistema propuesto, permitiendo complementar el trabajo realizado por los SIEM comerciales existentes.

Lo expuesto en este trabajo supone un pequeño extracto de la publicación original admitida y presentada en las JNIC 2019 (Bravo *et al.*, 2019), donde se puede obtener información más amplia y detallada sobre esta investigación realizada.

REFERENCIAS

Bravo, A., Sancho, J. y Caro, A. (2019). Extracción de conocimiento a partir de fuentes de datos reales procedentes de la monitorización de eventos de seguridad. Jornadas Nacionales de Investigación en Ciberseguridad (86-93). Cáceres, España: DEHESA, Repositorio institucional Universidad de Extremadura.

Fortinet. (2018). *Fortinet named a Leader in the 2018 Gartner Enterprise Firewall Magic Quadrant*. Obtenido de https://www.fortinet.com/products/next-generation-firewall.html

Internet Engineering Task Force, I. (2009). *The Syslog Protocol*. Obtenido de https://tools.ietf.org/html/rfc5424

Jouini, M., Rabai, L. y Aissa, A. (2014). Classification of security threats in information systems. 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014) (489-496). Hasselt, Bélgica: Elsevier.

Shostack, A. (2014). Threat Modeling: Designing for Security. Indianápolis, Estados Unidos: John Wiley & Sons, Inc.

Witten, I. y Frank, E. (2011). *Data Mining: Practical Machine Learning Tools and Techniques.* Second Edition. San Francisco, Estados Unidos: Elsevier.

APUNTES BIOGRÁFICOS

Alberto Bravo Gómez (Cáceres, 21 de noviembre de 1996) cursó el Grado en Ingeniería Informática en Ingeniería del Software en la Escuela Politécnica de Cáceres, para posteriormente, incorporarse a la Universidad de Extremadura como Personal de Apoyo a la Investigación, mientras cursaba el Máster Universitario en Ingeniería Informática impartido en la misma Escuela. En la actualidad reside en Cáceres

y se encuentra desarrollando investigaciones sobre los campos relacionados con el análisis de datos, modelos de aprendizaje automático y ciberseguridad aplicada a patrones de detección de amenazas y vulnerabilidades.

Contacto: **albertobg@unex.es**