Técnicas de machine Learning para la detección de Ransomware: Revisión sistemática de literatura

Machine Learning Techniques for Ransomware Detection: Systematic Literature Review

https://doi.org/10.5281/zenodo.7373655

AUTORES: **Oscar Miguel Cumbicus Pineda**[1]*

Pablo Vinicio Ludeña Preciado[2]

Lisset Alexandra Neyra Romero[3]

DIRECCIÓN PARA CORRESPONDENCIA: oscar.cumbicus@unl.edu.ec

Fecha de recepción: 19/ 06 / 2022

Fecha de aceptación: 10 / 09 / 2022

RESUMEN

El ransomware es uno de los problemas de seguridad informática más críticos, es un tipo de malware que cifra o bloquea la información de la víctima para solicitar el pago de un rescate y devolverles el acceso a sus datos. La presente investigación tuvo el propósito de identificar las técnicas y/o algoritmos de Machine Learning (ML) utilizadas para la detección y clasificación de las diferentes familias ransomware, así como las herramientas de software que se utilizan para la aplicación de estos algoritmos. Está revisión sistemática de literatura (RSL) se apoyó en la metodología propuesta por Bárbara Kitchenham y en el uso de la herramienta Parsifal. Los resultados obtenidos muestran que los algoritmos y/o técnicas de machine learning más utilizados son: Random Forest (RF) con el 23 %, Decisión Tree (DT) con un 14 %, Long Short-Term Memory (LSTM) utilizado en un 9 %, Support Vector Machine Learning (SVM) y Deep Neural Network (DNN) con el 6 %. Las herramientas más utilizadas para la aplicación de los algoritmos de machine learning,

[1]* Master en Ingeniería Computacional y Sistemas Inteligentes, Universidad Nacional de Loja, oscar.cumbicus@unl.edu.ec

[2]Ingeniero en Sistemas, Universidad Nacional de Loja

[3]Master en Ingeniería Computacional y Sistemas Inteligentes, Universidad Nacional de Loja

fueron Cuckoo Sandbox y Weka Framework con el 17 %. Llegando a la conclusión que el machine learning permite detectar en las etapas iniciales patrones de diferentes familias ransomware.

PALABRAS CLAVES:

Ransomware, Aprendizaje Automático, Aprendizaje Profundo, Algoritmos, Detección de Malware.

ABSTRAC

Ransomware is one of the most critical computer security problems; it is a type of malware that encrypts or blocks the victim's information to request the payment of a ransom and give them back access to their data. The present research had the purpose of identifying the techniques and/or Machine Learning (ML) algorithms used for the detection and classification of the different ransomware families, as well as the software tools used for the application of these algorithms. This systematic literature review (SLR) was based on the methodology proposed by Barbara Kitchenham and the use of the Parsifal tool. The results obtained show that the most commonly used machine learning algorithms and/or techniques are: Random Forest (RF) with 23 %, Decision Tree (DT) with 14 %, Long Short-Term Memory (LSTM) used with 9 %, Support Vector Machine Learning (SVM) and Deep Neural Network (DNN) with 6 %. The most used tools for the application of machine learning algorithms were Cuckoo Sandbox and Weka Framework with 17%. We conclude that machine learning allows us to detect patterns of the different ransomware families in the initial stages.

KEYWORDS

Ransomware, Machine Learning, Deep Learning, Algorithms, Malware Detection.

INTRODUCCIÓN

El mundo de hoy se ha vuelto dependiente del uso de sistemas informáticos como computadores, teléfonos inteligentes, entre otros, este tipo de dispositivos y la información que en ellos se almacena o es transmitida, se encuentra amenazada por ciberdelincuentes que buscan copiar, sustraer o atentar contra su integridad (Monje y Alexander, 2018).

Recientemente, se han producido muchos ataques cibernéticos que utilizan ransomware en varios sistemas en todo el mundo, y la cantidad de variantes ha aumentado rápidamente cada año. Los ataques no se limitan a personas particulares, sino también a organizaciones gubernamentales, bancarias, entre otras (Veloz et al., 2020).

La empresa ESET Threat reportó que la actividad del ransomware al final del segundo trimestre del año 2020, tuvo un aumento significativo de nuevos ataques ransomware, La variante llamada WannaCryptor obtuvo un 47,9 %, en comparación con el 40,5 % del primer trimestre del mismo año. Para el segundo trimestre del año 2020, apareció CryCryptor, un nuevo tipo de ransomware Android que se hizo pasar por una aplicación de rastreo de COVID-19 proporcionada por el gobierno de Canadá (Manzano et al., 2020). ESET también dio a conocer las variantes de ataques ransomware en los años 2016-2018, donde se observa que Ecuador ocupa el cuarto lugar con las variantes de este tipo de ataque.

El objetivo principal de la presente investigación fue identificar los algoritmos del Machine Learning para la detección y clasificación de las diferentes familias ransomware, así como las herramientas de software a utilizar para el entrenamiento del conjunto de datos con las cuales se puede identificar patrones, logrando así, brindar un apoyo en la detección de ransomware y de esta manera mitigar daños irreparables a la integridad de la información.

A continuación, se detalla el desarrollo de la presente investigación, el cual se realizó en las siguientes secciones: Metodología donde se describen las fases que propone Barbara Kitchenham, posteriormente en la sección de Resultados, se analiza la información extraída de los artículos seleccionados, luego se explica e interpreta de forma más específica la información en la sección de Discusión y finalmente, se especifica las Conclusiones obtenidas al elaborar la RSL.

METODOLOGÍA

A. Metodología para la revisión sistemática de literatura.

Para la planificación y conducción de la Revisión Sistemática de Literatura (RSL), se aplicó la metodología de Barbara Kitchenham cutas fases y tareas se describen a continuación:

Planificar la revisión sistemática de literatura con el uso de la herramienta Parsifal.

- Determinar las preguntas de investigación.
- Establecer el proceso de búsqueda.
- Definir los criterios de inclusión y exclusión para los artículos.
- Seleccionar las fuentes de consulta.
- Crear cadenas de búsqueda.
- Verificación de calidad.

Desarrollar la revisión sistemática de literatura con la planificación definida.

- Búsqueda de artículos.
- Aplicar la evaluación de la calidad a los artículos primarios escogidos.
- Selección de los artículos definitivos para el análisis de la información.
- Extracción de la información.
- Análisis y clasificación de la información.

Documentar e interpretar los resultados de la revisión.

- Desarrollar el informe del mapeo sistemático de la revisión.
- Desarrollar el informe de las preguntas de investigación de la revisión.

A continuación, presentamos las herramientas y recursos técnicos principalmente utilizados para el desarrollo de esta investigación.

B. Herramientas.

Zotero: Es una herramienta que permite la gestión y administración de fuentes bibliográficas; permitió gestionar todas las referencias mencionadas en este TT ayudando así al desarrollo de la SLR y ordenar la información de la misma (Corporation for Digital Scholarship, 2022).

Parsifal: Es una herramienta online que permitió el desarrollo de la SLR, ayudando así a automatizar y optimizar el tiempo que lleva el proceso de obtención y selección de información (Parsifal, 2022).

Thesaurus: Es una página web que permite encontrar los sinónimos de las palabras y el campo de la ciencia al cuál se orienta dicha palabra. Es utilizada para estructurar las cadenas de búsqueda en las bases de datos científicas seleccionadas (IEEE, 2022).

DISCUSIÓN Y RESULTADOS

A. Planificar la revisión sistemática de literatura.

A continuación, se detallan cada una de las tareas realizadas durante la etapa de planificación de a revisión sistemática de literatura:

Determinar las preguntas de investigación

Para establecer las preguntas de investigación en la SLR, se realizó una búsqueda exploratoria donde se analizó cuatro artículos al azar sobre el tema de estudio. Esta indagación permitió obtener un listado de palabras claves candidatas. Para finalizar se planteó seis preguntas de investigación clasificadas 3 de mapeo sistemático (MQ) y 3 de revisión sistemática (RQ):

Preguntas del Mapeo Sistemático (MQ).

MQ1. ¿Cuántos estudios se han publicado en los últimos cinco años acerca de las técnicas de machine learning para la detección de ransomware?

MQ2. ¿Cuáles son los autores más relevantes y activos en este ámbito de estudio?

MQ3. ¿En dónde se han publicado la mayor cantidad de los artículos sobre el tema de estudio?

Preguntas de la Revisión Sistemática (RQ).

RQ1. ¿Cuáles son los algoritmos del machine learning que mejores resultados obtuvieron para la detección de ataques ransomware?

RQ2. ¿Cómo ayuda el machine learning en la prevención de ataques ransomware?

RQ3. ¿Qué herramientas de software se han utilizado para la aplicación de técnicas y/o algoritmos de machine learning en la detección de ransomware?

Establecer el proceso de búsqueda

Se definieron los términos base aplicando el método PICOC propuesto por (Petticrew y Roberts, 2008) para definir el ámbito de la SLR. Sus componentes son la población, intervención, comparación resultados contextos. Este método permitió definir las expresiones que compusieron la cadena de búsqueda. Los términos clasificados se detallan a continuación:

Población (P): "Ransomware"

Intervención (I): "Machine Learning" OR "Deep Learning"

Comparación (C): No aplica

Resultados (O): "Algorithms" OR "Methods" OR "Techniques" OR "Classification"

Contexto (C): "Intelligence Artificial"

Clasificadas las palabras claves por medio de PICOC, se generó una tabla de palabras claves sinónimos que tienen relación con cada una de las categorías PICOC. La identificación de los términos sinónimos fue a través de la utilización del 2020 IEEE Thesaurus en (Cusack et al., 2018). En la Tabla 1 se detalla la lista de las palabras claves sinónimas.

Tabla 1: Definición de palabras clave

Palabra Clave	Sinónimos	Relación PICOC
Algorithms	Classification Detection Methods Techniques	Resultados
Machine Learning	Deep Learning	Intervención
Ransomware	s/n	Población

Establecer el proceso de búsqueda

Al obtener los resultados preliminares de las búsquedas y con el objetivo de filtrar la información relevante del tema de investigación, es conveniente describir el criterio a seguir para la selección de estudios, basados en la pregunta de investigación y según el protocolo

de Kitchenham (2004). Para incluir y excluir los estudios fue necesario establecer 4 criterios de inclusión (IC) y 5 criterios de exclusión (EC), los cuales se describen a continuación:

Criterios de inclusión (IC): En cuanto a los criterios de inclusión para la búsqueda se definieron los siguientes:

- **IC1**. Artículos publicados desde el 2017 en adelante.
- **IC2**. Artículos que contengan información de técnicas del machine learning y herramientas que se emplean para la detección de ransomware.
- **IC3**. Artículos escritos en idioma inglés.
- **IC4**. Artículos que hayan sido publicados en revistas científicas, artículos científicos y conferencias.

Criterios de exclusión (EC): En cuanto a los criterios de exclusión para la búsqueda se definieron los siguientes:

- **EC1**. Artículos duplicados.
- **EC2**. Artículos que no pertenecen al área Ciencias y Computación.
- **EC3**. Capítulos de libros, manuales, literatura gris
- **EC4**. Artículos donde el contenido sea similar a otros estudios quedándose solo estudio de contenido.
- **EC5**. Artículos cuyo título no tenga relación con el objeto de estudio.

Seleccionar las fuentes de consulta

Siguiendo los lineamientos de Kitchenham (2004), en la primera etapa se consideraron las fuentes de búsqueda indicadas en la Tabla 2, las cuales fueron seleccionadas basadas en la accesibilidad y admisión de consultas avanzadas.

Tabla 2. Bases de Datos Científicas

Base de Datos	URL
ACM	http://portal.acm.org

IEEE	http://ieeexplore.ieee.org
ScienceDirect	http://www.sciencedirect.com
Scopus	http://www.scopus.com

Cadenas de búsqueda

Se consideró palabras claves a partir de una revisión preliminar de artículos, mediante el uso de Thesaurus de IEEE se buscó sinónimos, se utilizó los operadores lógicos "AND/OR" con la finalidad de generar las cadenas de búsqueda.

La herramienta Parsifal generó la cadena de búsqueda general, la misma que se modificó con cada base de datos y todas las palabras claves se definieron en idioma inglés, mismas que se encuentran estructuradas de la siguiente manera, ver Tabla 3.

Tabla 3. Cadenas de búsqueda

Base de datos	Cadenas de Búsqueda
Parsifal	("Ransomware") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" OR "Classification" OR "Detection" OR "Methods" OR "Techniques")
ACM	(("Ransomware") AND ("Machine Learning" AND "Deep Learning") AND ("Algorithms" OR "Classification" OR "Detection" OR "Methods" OR "Techniques")) Publication Date: (01/01/2017 TO 12/31/2021)
IEEE	(("Ransomware") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" OR "Classification" OR "Detection" OR "Methods" OR "Techniques")) Filters Applied: 2017-2021

Science@Direct	(("Ransomware") AND ("Machine Learning" AND "Deep Learning") AND ("Algorithms" AND "Classification" OR "Detection" AND ("Methods" OR "Techniques")) Year: 2017-2021 Type: Research articles
Scopus	TITLE-ABS-KEY(("Ransomware") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" AND ("Classification" OR "Detection") AND ("Methods" OR "Techniques"))) AND (LIMIT-TO (DOCTYPE,"ar") OR LIMIT-TO (DOCTYPE,"cp")) AND (LIMIT-TO (SUBJAREA,"COMP")) AND (LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019) OR LIMIT-TO (PUBYEAR,2018) OR LIMIT-TO (PUBYEAR,2017))

Verificación de calidad

Definida la cadena de búsqueda es imprescindible realizar la evaluación de calidad de los documentos seleccionados. Cada artículo es evaluado siguiendo los criterios de la Base de Datos de Resúmenes de Revisiones de Efectos (DARE) del centro de revisiones de Difusión (CRD) de la Universidad de York, según lo explicado por (Brereton *et al.*, 2007). Las siguientes preguntas se establecieron para evaluar la calidad de los artículos preseleccionados:

QA1. ¿El autor cita o utiliza alguna técnica y/o algoritmo del machine learning para la detección de ataques ransomware?

QA2. ¿En los estudios se menciona alguna herramienta de software utilizada para la aplicación de técnicas y/o algoritmos para la detección de ransomware?

QA3. ¿En los artículos se habla acerca de la importancia del machine learning para la detección de ataques ransomware?

Para finalizar se determinaron los parámetros de puntuación para definir que artículos serán seleccionados y rechazados. Los parámetros se detallan a continuación:

- Si la respuesta es Si su puntuación será 1.0
- Si la respuesta es Parcialmente su puntuación será 0.5
- Si la respuesta es No su puntuación será 0.0

B. Desarrollar la revisión sistemática de literatura.

Búquedas de artículos

Se implementó la cadena de búsqueda en cada una de las bases de datos seleccionadas (ACM, IEEE, ScienceDirect y Scopus), donde se obtuvieron 418 artículos. Por cada búsqueda que se realizó se generó un archivo con extensión bib, el mismo que se cargó en la herramienta Parsifal.

De los 418 artículos, se identificaron y eliminaron 35 artículos duplicados, quedando 383 artículos por revisar.

Evaluación de calidad a los artículos primarios

Los 383 artículos obtenidos en la sección anterior fueron revisados y analizados en su título y resumen, tomando en consideración los criterios de inclusión y exclusión. Del total se eliminaron 312 artículos que son irrelevantes al objeto de estudio, además se descartaron porque su argumentación referente a técnicas de machine learning para la detección de ransomware es débil y no dan contestación a las preguntas de investigación planteadas. Obteniendo 71 artículos para su revisión y análisis.

Selección de artículos definitivos

Se aplicaron las preguntas de calidad a cada uno de los 71 artículos seleccionados, y se seleccionaron 34 los cuales tenian una calificación igual o superior a 2 puntos. En la Tabla 4 se puede observar la información relevante de cada uno de los artículos seleccionados.

Tabla 4. Artículos seleccionados

Código	Título	Referencia
SA01	A Novel Malware Analysis for Malwarem Detection and Classification Using Machine Learning Algorithms	(Sethi et al., 2017)
SA02	A Survey on Malware Detection with Deep Learning.	(Sahin y Bahtiyar, 2020)
SA03	Analysis of Machine Learning Techniques for Ransomware Detection.	(Noorbehbahani et al., 2019)
SA04	An Empirical Comparison of Supervised Algorithms for Ransomware Identification on Network Traffic.	(Manzano et al., 2020)
SA05	Attention in Recurrent Neural Networks for Ransomware Detection.	(Agrawal et al., 2019)
SA06	Android Ransomware Detection using Machine Learning Techniques: A Comparative Analysis on GPU and CPU.	(S. Sharma et al., 2020)
SA07	A Digital DNA Sequencing Engine for Ransomware Detection Using Machine Learning.	(Khan et al., 2020)
SA08	An Efficient Machine Learning-based Approach for Android v.11 Ransomware Detection.	(Almomani et al., 2021)

SA09	Analysing Indicator of Compromises for Ransomware: Leveraging IOCs with Machine Learning Techniques.	(Verma et al., 2018)
SA10	A review on android ransomware detection using deep learning techniques.	(Alzahrani y Alghazzawi, 2019)
SA11	Behavioral-Based Classification and Identification of Ransomware Variants Using Machine Learning.	(Daku et al., 2018)
SA12	Bidirectional long short-term memory classifier assist for intelligent ransomware detection in Android OS.	(Siłka, 2021)
SA13	Classification of ransomware families with machine learning based onN-gram of opcodes.	(Zhang et al., 2019)
SA14	Detecting Ransomware Using Support Vector Machines.	(Takeuchi et al., 2018)
SA15	Deep Learning for Detecting Ransomware in Edge Computing Devices Based on Autoencoder Classifier.	(AbdulsalamYa'u et al., 2019)
SA16	Deep learning LSTM based ransomware detection.	(Maniath et al., 2017)
SA17	Detecting Android Locker-Ransomware on Chinese Social Networks.	(Su et al., 2019)
SA18	Detecting Ransomware Using Process Behavior Analysis.	(Arabo et al., 2020)
SA19	Detecting ransomware attacks using intelligent algorithms:recent development and next direction from deep learning and big data perspectives.	(Bello et al., 2021)

SA20	Exposing Android Ransomware Using Machine Learning.	(Victoriano, 2019)
SA21	Feature-Selection Based Ransomware Detection with Machine Learning of Data Analysis.	(Wan et al., 2018)
SA22	Identification of Ransomware families by Analyzing Network Traffic Using Machine Learning Techniques.	(Almousa et al., 2021)
SA23	Leveraging Deep Learning Models for Ransomware Detection in the Industrial Internet of Things Environment.	(Al-Hawawreh y Sitnikova, 2019)
SA24	Machine Learning Based Detection of Ransomware Using SDN.	(Cusack et al., 2018)
SA25	Machine Learning Based Ransomware Detection Using Storage Access Patterns Obtained. From Live-forensic Hypervisor	(Hirano y Kobayashi, 2019)
SA26	On the classification of MicrosoftWindows ransomware using hardware profile.	(Aurangzeb et al., 2021)
SA27	Ransomware Detection with Semi Supervised Learning.	(Noorbehbahani y Saberi, 2020)
SA28	Ransomware Noise Identification and Eviction Through Machine Learning Fundamental Filters.	(P. Sharma et al., 2019)
SA29	Ransomware Detection Using Deep Learning in the SCADA System of Electric Vehicle Charging Station.	(Basnet et al., 2021)
SA30	Ransomware Detection in Executable Files Using Machine Learning.	(Ganta et al., 2020)
SA31	Ransomware Detection Using Limited Precision Deep Learning Structure in FPGA.	(Alrawashdeh y Purdy, 2018)

SA32	RansomWall: A layered defense system against cryptographic ransomware attacks using machine learning.	(Shaukat y Ribeiro, 2018)
SA33	Ransomware: Let's fight back!. Using Machine Learning Algorithms	(Chadha y Kumar, 2017)
SA34	RansomDroid: Forensic analysis and detection of Android Ransomware using unsupervised machine learning technique.	(S. Sharma et al., 2021)

Extracción de la información

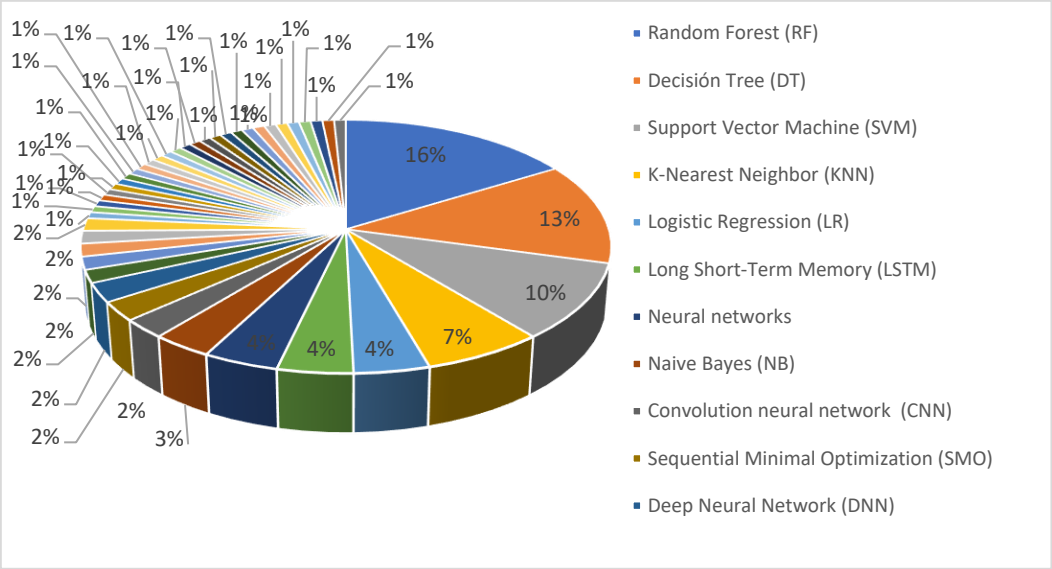
Para el desarrollo de esta fase se, revisó y analizó la información de cada uno de los artículos seleccionados para el análisis, en el cual se detalla sus principales técnicas y/o algoritmos y herramientas. Se tomaron como pautas para la extracción y el análisis de la información lo siguiente:

- Tomar como base de estudio los documentos que mencionen o analicen el mismo tipo de información ya sea de técnicas y/o algoritmos y herramientas.
- Al existir una gran cantidad de técnicas propuestas por los autores se procede a clasificarlas en un solo apartado ya que cada autor adapta su proceso de machine learning a un caso específico para obtener los resultados esperados.

Análisis y clasificación de la información

Análisis de algoritmos: Se tomó en cuenta el número total de artículos donde se menciona los algoritmos utilizados, estos datos se los representó en un grado porcentual para hacer visible de una mejor manera la información en la Figura 1.

Figura 1: Algoritmos utilizados

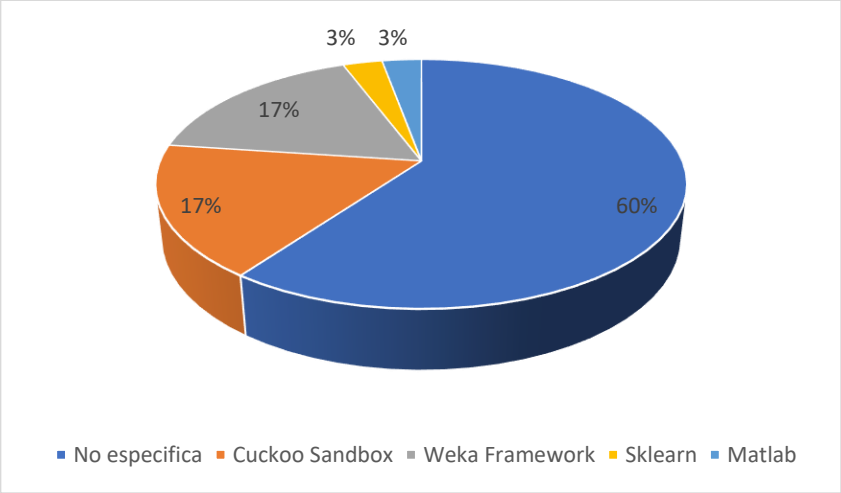


Al observar y analizar la Figura 1 se puede determinar lo siguiente:

- El algoritmo Random Forest (RF) con el 16% obtiene un porcentaje mayor en el gráfico, debido a que engloba a varios artículos donde se utiliza este algoritmo como técnica del machine learning para la detección de ransomware.
- Decisión Tree (DT) obtiene el 13% de usabilidad; Máquina de Vectores de Soporte (SVM) es usado en el 10% de los artículos; El algoritmo K-Nearest Neighbor (KNN) es utilizado en el 7% de los artículos analizados.
- El 4%, es alcanzado por Logistic Regression (LR), Long Short-Term Memory (LSTM) y Neural Networks (NN).
- Naive Bayes (NB), obtiene el 3% de los artículos analizados.
- Convolution Neural Network (CNN), Deep Neural Network (DNN), Sequential Minimal Optimization (SMO), son utilizados en un porcentaje del 2% dentro de los artículos seleccionados.
- Los algoritmos de Recurrent Neural Network (RNN), Gain Ratio, One R, Quadratic Discriminant Analysis (QDA), Adabost, Random Tree, entre otros son utilizados en un porcentaje mínimo de artículos siendo referenciados en un solo artículo de investigación.

Análisis de herramientas: Se tomó en cuenta el número total de artículos donde se menciona los métodos de extracción, estos datos se los represento en un grado porcentual para hacer visible de una mejor manera la información en la Figura 2.

Figura 2: Herramientas utilizadas



Al observar y analizar la Figura 2 se puede determinar lo siguiente:

- En el 60% de los artículos no se especifíca que tipo de herramientas son utilizadas en el proceso de aplicación de las técnicas de machine learning para la identificación de ataques de ransomware.
- Las herramientas Cuckoo Sandbox y Weka Framework son utilizadas en el (17%) de artículos para la aplicación de las técnicas y/o algoritmos.
- Las herramientas Sklearn y Matlab son utilizadas en un porcentaje del (3%) del total de los artículos analizados.

C. Documentar e interpretar los resultados de la revisión.

Esta fase comprende el análisis, la interpretación de la información obtenida y la presentación de los resultados, en donde se da contestación a las preguntas de investigación tanto de mapeo sistemático (MQ) y de la revisión sistemática (RQ), definidas en la fase de planificación.

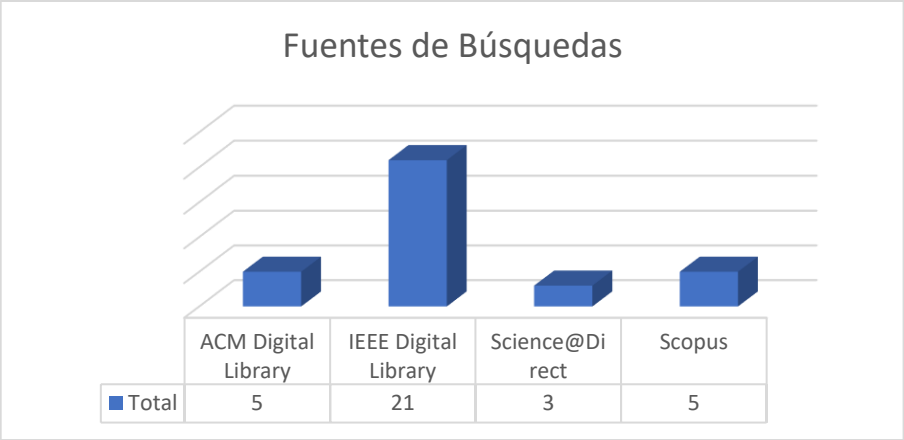
Informe del mapeo sistemático

En el siguiete aparatado se dan a conocer las respuestas de las preguntas de mapeo sistemático (MQ).

MQ1. *¿Cuántos estudios se han publicado en los últimos cinco años acerca de las técnicas de machine learning para la detección de ransomware?*

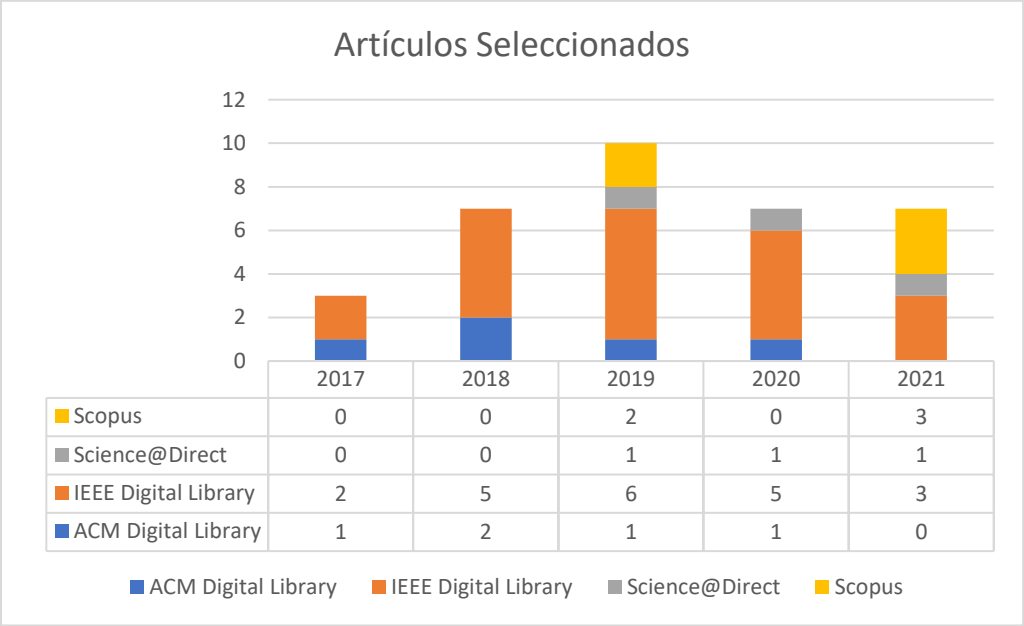
En cuanto a MQ1, tomando en cuenta los últimos cinco años se han publicado 34 artículos, siendo IEEE Digital Library el mayor aportador de publicaciones científicas con 21 aportaciones, seguido de ACM Digital Library y Scopus con 5 aportaciones cada uno respectivamente y Science direct con 3 aportaciones, se ilustra en la Figura 3.

Figura 3: Artículos seleccionados de los motores de búsqueda



En el año 2019, se publicaron la mayor cantidad de publicaciones de artículos, y la mayor parte de esas aportaciones científicas se encuentran en bases de datos como IEEE Digital Library con 6 aportaciones en la Figura 4 podemos ver detalladamente el progreso de la publicaciones en este ámbito de estudio.

Figura 4: Artículos publicados por años



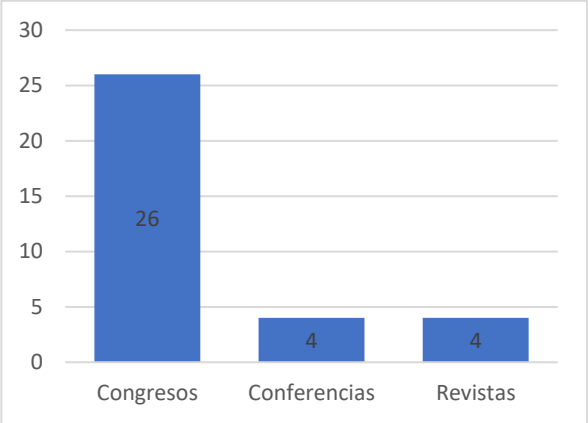
MQ2. *¿Cuáles son los autores más relevantes y activos en este ámbito de estudio?*

Con respecto a MQ2 se identifican los autores más releventes que aportan en investigaciones para la detección de ransomware a través de las técnicas del machine learning. Los autores se identificaron a partir de los trabajos seleccionados. Dando como resultado que los Autores: Noorbehbahani, Fakhroddin; Saberi, Mohammad; Sharma, Shweta; Krishna, C. Rama; Kumar, Rakesh publicaron 2 artículos relacionados al tema de estudio.

MQ3. *¿En dónde se han publicado la mayor cantidad de los artículos sobre el tema de estudio?*

Con respecto a MQ3 en la Figura 5, se muestra en detalle la cantidad de artículos publicados donde obtenemos 26 artículos en congresos y 4 en conferencias y revistas.

Figura 5: Tipo de publicaciones



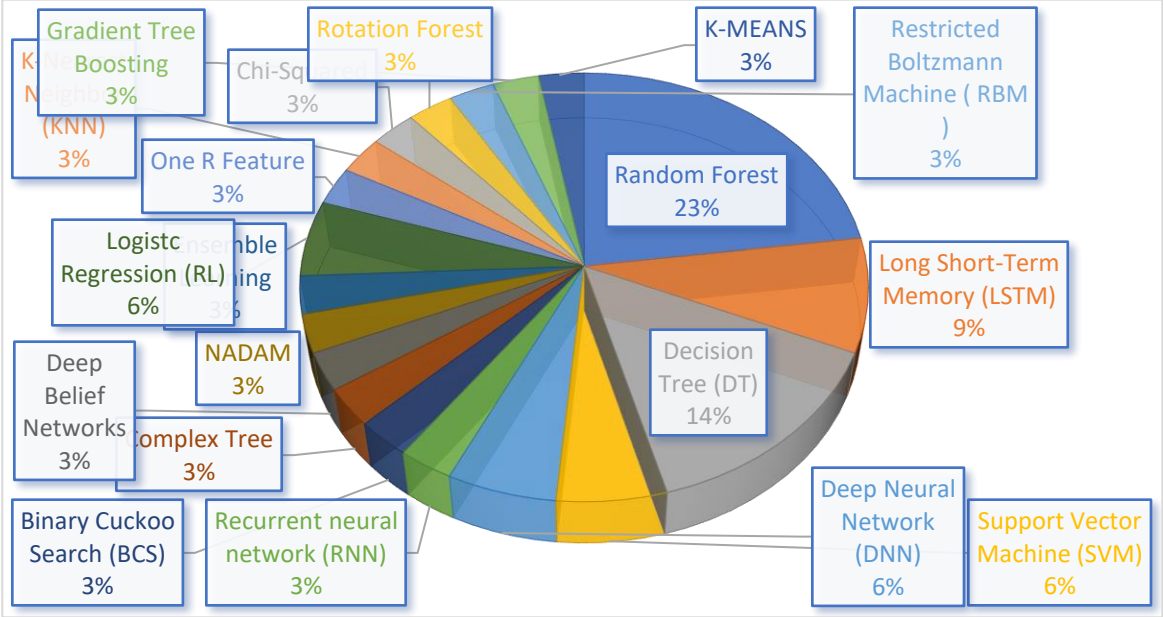
Análisis de la revisión sistemática

RQ1. *¿Cuáles son los algoritmos del machine learning que mejores resultados obtuvieron para la detección de ataques ransomware?*

Para poder dar contestacion a RQ1, se revisó y analizó los 34 artículos seleccionados, donde se identificó 19 algoritmos y/o técnicas de machine learning utilizados para la detección de ataques ransomware. De cada artículo se extrajo el algoritmo que mejor rendimiento aporto al investigador.

En el análisis se identificó que las técnicas y/o algoritmos que proporcionan mejores rendimientos en la detección de las diferentes familias de ransomware fue Random Forest (RF), obteniendo un resultado del 23%; Decision Tree (DT), utilizado en SA01, SA11, SA19, SA21 y SA23 obtuvo un valor porcentual del 14%; Long Short-Term Memory (LSTM) utilizado en SA05, SA16, SA30 con un 9%; (SVM) utilizado en SA14 y SA34, Deep Neural Network (DNN) utilizado en SA15 y SA24, obtubieron el 6% de usabilidad. En la Figura 6, se muestra a detalle los resultados encontrados.

Figura 6: Principales Técnicas y/o algoritmos encontrados



En conlusión de los 34 artículos analizados apreciamos que las técnicas que mejores prestaciones brindó a los investigadores para el análisis y detección del ransomware fue aplicando Random Forest (RF) con el 23%, Decision Tree (DT) con 14% y Long Short-Term Memory (LSTM) con un 9%, los mismos que fueron mencionados en la mayor parte de los estudios primarios.

RQ2. *¿Ayuda el machine learning en la prevención de ataques ransomware?*

De acuerdo SA04, estudiar los métodos de machine learning permite la clasificación e identificación de nuevos casos de ransomware, dado que este enfoque no depende de las características predefinidas, sino que las construye internamente como parte del proceso de aprendizaje. En el artículo SA08 se menciona que, ejecutar algoritmos de machine learning

permite generar modelos predictivos de ransomware con alta precisión y eficiencia. SA26 menciona que el machine learning automatiza el análisis de ataques ransomware mediante la utlización de técnicas de aprendizaje automático, dado que permiten obtener patrones de ransomware. En conclusión el machine learning ayuda a la prevención de ataques ransomware prediciendo las amenazas antes que ocurran, esto se logra en base al entrenamiento utilizando algoritmos para encontrar patrones en altas cantidades de información lo cual permite clasificar y reconocer el ataque en sus etapas iniciales y así poder tomar las medidas correctoras.

RQ3. *¿Qué herramientas de software se han utilizado para el análisis de técnicas de machine learning en la detección de ransomware?*

En los artículos seleccionados se identificaron los siguientes resultados para dar contestación a RQ3: el 60% de los artículos analizados, no especifican el tipo de herramienta que utilizan para la aplicación de las técnicas de machine learning para la detección de ransomware, el 17% de los artículos mencionan las herramientas Cuckoo Sandbox y Weka Framework para ejecutar los algoritmos de Machine Learning que permiten la detección de las diferentes familias de ransomware y en un menor porcentaje del 3% de los artículos utilizan las herramientas Matlab y Sklearn. En conclusión con la extracción de la información en los estudios realizados se determinó que las herramienta más utilizada es Cuckoo Sandbox y Weka Framework, pero en sí, no podemos decir que una herramienta es mejor que otra ya que depende de las necesidades del investigador y sobre los conocimientos que tenga al utilizar dicha herramienta para el análisis y obtención de los resultados requeridos.

DISCUSIÓN

La extracción de la información facilitó responder las preguntas de mapeo (MQ), obteniendo como resultado que el mayor aportador de estudios en este trabajo es la fuente de consulta IEEE, y en el año 2019 se encuentra la mayor cantidad de artículos publicados en los últimos cinco años. Además, se determinó cuales son los autores más relevantes que aportan activamente en investigaciones para la detección de ransomware a través del uso de técnicas del machine learning. Después se procedió a dar contestación a las preguntas de la revisión

(RQ), las mismas que se obtuvieron realizando la interpretación de los resultados obtenidos donde se analizó lo siguiente: Entre las técnicas y/o algoritmos más usados en los artículos revisados tenemos: Random Forest (RF) el mismo que brindó mejores resultados en la mayor parte de artículos seleccionados, seguido de Decision Tree (DT) y Long Short-Term Memory (LSTM), además, cabe destacar que algoritmos como Recurrent neural network (RNN), Neural Networks (NN), Binary Cuckoo Search (BCS), Complex Tree, K-Nearest Neighbor (KNN), Gradient Tree Boosting, K-MEANS entre otros algoritmos son utilizados en un solo artículo. En base a la investigación realizada se determinó que al utilizar técnicas del Machine learning ayudaría en el ámbito de la seguridad informática para la detección proactiva de ataques ransomware, así como el descubrimiento asistido de características o patrones generados por este tipo de malware para vulnerar sistemas informáticos. Igualmente, se identificó que Cuckoo Sandbox y Weka Framework como las herramientas más utilizadas para la aplicación de las técnicas y/algoritmos de machine learning para la detección de ransomware. Para finalizar respondiendo la pregunta del problema de investigación: "¿Cuáles son las técnicas de machine learning para la detección de ransomware?". En base a la información obtenida del análisis de los artículos se determinó que el algoritmo que mejores resultados brinda a los investigadores es Random Forest (RF) dado que es una buena solución para la detección y clasificación familias de ransomware convirtiéndolo en uno de los algoritmos con mejores resultados en comparación con otros.

CONCLUSIÓN

El uso de la herramienta Parsifal permitió el desarrollo óptimo y eficiente del proceso de la planificación de la revisión sistemática de literatura, lo que ayudó a gestionar la investigación de manera ordenada, siguiendo una serie de pasos precisos y específicos para extraer la información de los artículos relevantes que cumplieron con los criterios de inclusión y exclusión definidos con el fin de identificar las técnicas del machine learning para la detección de ransomware.

De los 34 artículos que fueron analizados se obtuvo que las técnicas más utilizadas para la detección de ransomware son: el algoritmo Random Forest (RF) citado en la mayor parte de artículos seleccionados como el que mejor rendimiento obtiene en la detección y clasificación de ransomware, seguido de Decision Tree (DT) y Long Short-Term Memory (LSTM). Las

herramientas de software como Cuckoo Sandbox y Weka Framework fueron las más utilizadas para la aplicación de las técnicas y/o algoritmo para identificar y clasificar las diferentes familias ransomware.

El uso técnicas basadas en machine learning son de gran ayuda, permiten identificar patrones para evitar ataques ransomware y mitigar daños irreparables de pérdidas de información.

REFERENCIAS BIBLIOGRÁFICAS

AbdulsalamYa'u, G., Job, G. K., Waziri, S. M., Jaafar, B., SabonGari, N. A., y Yakubu, I. Z. (2019, diciembre). Deep Learning for Detecting Ransomware in Edge Computing Devices Based On Autoencoder Classifier. En 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT) (pp. 240–243). Mysuru, India: IEEE. Descargado 2022-01-23, de https:// ieeexplore.ieee.org/document/9114576/ doi: 10 .1109/ICEECCOT46775.2019.9114576

Agrawal, R., Stokes, J. W., Selvaraj, K., y Marinescu, M. (2019, mayo). Attention in Recurrent Neural Networks for Ransomware Detection. En ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3222–3226). Brighton, United Kingdom: IEEE. Descargado 2022-01-23, de https://ieeexplore.ieee.org/document/8682899/ doi: 10.1109/ICASSP.2019.8682899

Al-Hawawreh, M., y Sitnikova, E. (2019, noviembre). Leveraging Deep Learning Models for Ransomware Detection in the Industrial Internet of Things Environment. En 2019 Military Communications and Information Systems Conference (MilCIS) (pp. 1–6). Canberra, Australia: IEEE. Descargado 2022-01-23, de https://ieeexplore.ieee .org/document/8930732/ doi: 10.1109/MilCIS.2019 .8930732

Almomani, I., AlKhayer, A., y Ahmed, M. (2021, abril). An Efficient Machine Learning-based Approach for Android v.11 Ransomware Detection. En 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA) (pp. 240–244). Riyadh, Saudi Arabia: IEEE.

Descargado 2022-01-23, de https://ieeexplore.ieee .org/document/9425059/ doi: 10.1109/CAIDA51941 .2021.9425059

Almousa, M., Osawere, J., y Anwar, M. (2021, septiembre). Identification of Ransomware families by Analyzing Network Traffic Using Machine Learning Techniques. En 2021 Third International Conference on Transdisciplinary AI (TransAI) (pp. 19–24). doi: 10.1109/TransAI51903 .2021.00012

Alrawashdeh, K., y Purdy, C. (2018, julio). Ransomware Detection Using Limited Precision Deep Learning Structure in FPGA. En NAECON 2018 - IEEE National Aerospace and Electronics Conference (pp. 152–157). Dayton, OH, USA: IEEE. Descargado 2022-01-23, de https://ieeexplore.ieee.org/document/8556824/ doi: 10.1109/NAECON.2018.8556824

Alzahrani, N., y Alghazzawi, D. (2019, noviembre). A

Review on Android Ransomware Detection Using Deep Learning Techniques. En Proceedings of the 11th International Conference on Management of Digital EcoSystems (pp. 330–335). Limassol Cyprus: ACM. Descargado: 2022-01-23, de https://dl.acm.org/doi/10.1145/3297662.3365785 doi: 10.1145/3297662.3365785

Arabo, A., Dijoux, R., Poulain, T., y Chevalier, G. (2020). Detecting Ransomware Using Process Behavior Analysis. 2022-01-23, de https://linkinghub.elsevier.com/ retrieve/pii/S1877050920303884 doi: 10 .1016/j.procs.2020.02.249

Aurangzeb, S., Bin Rais, R., Aleem, M., Islam, M., y Iqbal, M. (2021). On the classification of MicrosoftWindows ransomware using hardware profile. PeerJ Computer Science, 7, 1–24. doi: 10.7717/peerj-cs.361

Basnet, M., Poudyal, S., Ali, M. H., y Dasgupta, D. (2021, septiembre). Ransomware Detection Using Deep Learning in the SCADA System of Electric Vehicle Charging Station. En 2021 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America) (pp. 1–5). (ISSN: 2643-8798) doi: 10.1109/ ISGTLatinAmerica52371.2021.9543031

Bello, I., Chiroma, H., Abdullahi, U. A., Gital, A. Y., Jauro, F., Khan, A., ... Abdulhamid, S. M. (2021, septiembre). Detecting ransomware attacks using intelligent algorithms: recent development and next direction from deep learning and big data perspectives. Journal of Ambient Intelligence and Humanized Computing, 12(9), 8699–8717.

Descargado 2022-01-23, de https://link.springer.com/10.1007/s12652-020-02630-7 doi: 10.1007/s12652-020-02630-7

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., y Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. Journal of systems and software, 80(4), 571–583.

Chadha, S., y Kumar, U. (2017, mayo). Ransomware: Let's fight back! En 2017 International Conference on Computing, Communication and Automation (ICCCA) (pp. 925– 930). Greater Noida: IEEE. Descargado 2022-01-23, de http://ieeexplore.ieee.org/document/8229926/ doi: 10.1109/CCAA.2017.8229926

Corporation for Digital Scholarship. (2022, 01 03). Zotero. Retrieved from https://www.zotero.org/

Cusack, G., Michel, O., y Keller, E. (2018, marzo). Machine Learning-Based Detection of Ransomware Using SDN. En Proceedings of the 2018 ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization (pp. 1–6). Tempe AZ USA: ACM. Descargado 2021-12-10, de https://dl.acm.org/doi/10.1145/3180465.3180467 doi: 10.1145/3180465.3180467

Daku, H., Zavarsky, P., y Malik, Y. (2018, agosto).

Behavioral-Based Classification and Identification of Ransomware Variants Using Machine Learning. En 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (pp. 1560–1564). New York, NY, USA: IEEE. Descargado 2022-01-23, de https://ieeexplore.ieee.org/document/8456093/ doi: 10.1109/TrustCom/BigDataSE.2018.00224

Ganta, V. G., Harish, G., Kumar, V., y Rao, G. K. (2020, noviembre). Ransomware Detection in Executable Files Using Machine Learning. En 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) (pp. 282– 286). Bangalore, India: IEEE. Descargado 2022-01-23, de https://ieeexplore.ieee.org/document/9315672/ doi: 10.1109/RTEICT49044.2020.9315672

Hirano, M., y Kobayashi, R. (2019, octubre). Machine Learning Based Ransomware Detection Using Storage Access Patterns Obtained From Live-forensic Hypervisor. En 2019 Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS) (pp. 1– 6). Granada, Spain: IEEE. Descargado 2022-01-23, de https://ieeexplore.ieee.org/document/8939214/ doi: 10.1109/IOTSMS48152.2019.8939214

IEEE. (2022). IEEE. Retrieved from https://www.ieee.org/publications/services/thesaurus.html

Khan, F., Ncube, C., Ramasamy, L. K., Kadry, S., y Nam, Y. (2020). A Digital DNA Sequencing Engine for Ransomware Detection Using Machine Learning. IEEE Access, 8, 119710–119719. Descargado 2022-01-23, de https://ieeexplore.ieee.org/document/9121260/ doi: 10.1109/ACCESS.2020.3003785

Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. , 33.

Maniath, S., Ashok, A., Poornachandran, P., Sujadevi, V., Sankar A.U., P., y Jan, S. (2017, octubre). Deep learning LSTM based ransomware detection. En 2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE) (pp. 442–446). Noida: IEEE. Descargado 2022-01-23, de https://ieeexplore.ieee.org/ document/8358312/ doi: 10 .1109/RDCAPE .2017.8358312

Manzano, C., Meneses, C., y Leger, P. (2020, noviembre). An Empirical Comparison of Supervised Algorithms for Ransomware Identification on Network Traffic. En 2020 39th International Conference of the Chilean Computer Science Society (SCCC) (pp. 1–7). Coquimbo, Chile: IEEE. Descargado 2022-01-23, de https:// ieeexplore.ieee.org/document/9281283/ doi: 10 .1109/SCCC51225.2020.9281283

Monje, G., y Alexander, R. (2018). SEGURIDAD INFORMÁTICA Y EL MALWARE. , 11.

Noorbehbahani, F., Rasouli, F., y Saberi, M. (2019). Analysis of machine learning techniques for ransomware detection. En 2019 16th international isc (iranian society of cryptology) conference on information security and cryptology (iscisc) (pp. 128–133).

Noorbehbahani, F., y Saberi, M. (2020, octubre). Ransomware Detection with Semi-Supervised Learning. En 2020 10th International Conference on Computer and

Knowledge Engineering (ICCKE) (pp. 024–029). Mashhad, Iran: IEEE. Descargado 2022-01-23, de https:// ieeexplore.ieee.org/document/9303689/ doi: 10 .1109/ICCKE50421.2020.9303689

Parsifal. (2022). Parsifal. Retrieved from https://parsif.al/

Petticrew, M., y Roberts, H. (2008). Systematic Reviews in the Social Sciences: A Practical Guide. John Wiley & Sons. (Google-Books-ID: ZwZ1_xU3E80C)

Sahin, M., y Bahtiyar, S. (2020, noviembre). A Survey on Malware Detection with Deep Learning. En 13th International Conference on Security of Information and Networks (pp. 1–6). New York, NY, USA: Association for Computing Machinery. Descargado 2022-01-23, de https://doi.org/10.1145/3433174.3433609 doi: 10.1145/3433174.3433609

Sethi, K., Chaudhary, S. K., Tripathy, B. K., y Bera, P. (2017, octubre). A novel malware analysis for malware detection and classification using machine learning algorithms. En Proceedings of the 10th International Conference on Security of Information and Networks (pp. 107– 113). New York, NY, USA: Association for Computing Machinery. Descargado 2022-01-23, de https:// doi.org/10.1145/3136825.3136883 doi: 10.1145/ 3136825.3136883

Sharma, P., Chaudhary, K., Khan, M., y Wagner, M. (2019, diciembre). Ransomware Noise Identification and Eviction Through Machine Learning Fundamental Filters. En 2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (pp. 1–8). Melbourne, Australia: IEEE. Descargado 2022-01-23, de https:// ieeexplore.ieee.org/document/9162376/ doi: 10.1109/CSDE48274.2019.9162376

Sharma, S., Krishna, C. R., y Kumar, R. (2020, noviembre). Android Ransomware Detection using Machine Learning Techniques: A Comparative Analysis on GPU and CPU. En 2020 21st International Arab Conference on Information Technology (ACIT) (pp. 1–6). Giza, Egypt: IEEE. Descargado 2022-01-23, de https://ieeexplore.ieee.org/document/9300108/ doi: 10.1109/ACIT50332 .2020.9300108

Sharma, S., Krishna, C. R., y Kumar, R. (2021, junio). RansomDroid: Forensic analysis and detection of Android Ransomware using unsupervised machine learning technique.

Forensic Science International: Digital Investigation, 37, 301168. Descargado 2022-01-23, de https://linkinghub.elsevier.com/retrieve/pii/S2666281721000767 doi: 10.1016/j.fsidi.2021.301168

Shaukat, S. K., y Ribeiro, V. J. (2018, enero). RansomWall: A layered defense system against cryptographic ransomware attacks using machine learning. En 2018 10th International Conference on Communication Systems & Networks (COMSNETS) (pp. 356–363). Bengaluru: IEEE. Descargado 2022-01-23, de http://ieeexplore.ieee.org/document/8328219/ doi: 10.1109/COMSNETS .2018.8328219

Siłka, J. (2021). Bidirectional long short-term memory classifier assist for intelligent ransomware detection in Android OS. En (Vol. 2915, pp. 45–53). (ISSN: 1613-0073)

Su, D., Liu, J., Wang, X., y Wang, W. (2019). Detecting Android Locker-Ransomware on Chinese Social Networks. IEEE Access, 7, 20381–20393. Descargado 2022-01-23, de https://ieeexplore.ieee.org/document/8580446/ doi: 10.1109/ACCESS.2018.2888568

Takeuchi, Y., Sakai, K., y Fukumoto, S. (2018, agosto). Detecting Ransomware using Support Vector Machines. En Proceedings of the 47th International Conference on Parallel Processing Companion (pp. 1–6). Eugene OR USA: ACM. Descargado 2022-01-23, de https:// dl.acm.org/doi/10.1145/3229710.3229726 doi: 10.1145/3229710.3229726

Veloz, F. D. B., López, L. I. B., Valdivieso, L., y Álvarez, M. B. H. (2020). Indicadores para la detección de ataques. , 15.

Verma, M., Kumarguru, P., Brata Deb, S., y Gupta, A. (2018, noviembre). Analysing Indicator of Compromises for Ransomware: Leveraging IOCs with Machine Learning Techniques. En 2018 IEEE International Conference on Intelligence and Security Informatics (ISI) (pp. 154– 159). Miami, FL: IEEE. Descargado 2022-01-23, de https://ieeexplore.ieee.org/document/8587409/ doi: 10.1109/ISI.2018.8587409

Victoriano, O. B. (2019, octubre). Exposing Android Ransomware using Machine Learning. En Proceedings of the 2019 International Conference on Information System and System Management (pp. 32–37). Rabat Morocco: ACM. Descargado 2022-01-23, de https://dl.acm.org/doi/10.1145/3394788.3394923 doi: 10.1145/3394788 .3394923

Wan, Y.-L., Chang, J.-C., Chen, R.-J., y Wang, S.-J. (2018, abril). Feature-Selection-Based Ransomware Detection with Machine Learning of Data Analysis. En 2018 3rd International Conference on Computer and Communication Systems (ICCCS) (pp. 85–88). Nagoya, Japan: IEEE. Descargado 2022-01-23, de https://ieeexplore.ieee .org/document/8463300/ doi: 10.1109/CCOMS.2018 .8463300

Zhang, H., Xiao, X., Mercaldo, F., Ni, S., Martinelli, F., y Sangaiah, A. K. (2019, enero). Classification of ransomware families with machine learning based on N -gram of opcodes. Future Generation Computer Systems, 90, 211– 221. Descargado 2022-01-23, de https://linkinghub.elsevier.com/retrieve/pii/S0167739X18307325doi:10.1016/j.future.2018.07.052