

Valérie Gauthier-Umaña
Rafael Alberto Méndez-Romero
Jeimy Cano
Jorge Ramió Aguirre
Luis Enrique Sánchez Crespo
—Editores académicos—

Seguridad Informática

X Congreso Iberoamericano

CIBSI 2020



Universidad del
Rosario



MACC
Matemáticas Aplicadas y
Ciencias de la Computación



Congreso Iberoamericano de
Seguridad Informática 2019

SEGURIDAD INFORMÁTICA

X Congreso Iberoamericano, CIBSI 2020

Valérie Gauthier-Umaña
Rafael Alberto Méndez-Romero
Jeimy Cano
Jorge Ramió Aguirre
Luis Enrique Sánchez Crespo
—*Editores académicos*—



Universidad del
Rosario



MACC
Matemáticas Aplicadas y
Ciencias de la Computación



Congreso Iberoamericano de
Seguridad Informática 2019

Bogotá D. C., 2020

Editorial Universidad del Rosario
Universidad del Rosario

© Varios autores

ISBN-e: 978-958-784-393-4

<https://doi.org/10.12804/si9789587844337>

Diseño portada y preliminares
Precolombi EU, David Reyes

Hecho en Colombia
Made in Colombia

Los conceptos y opiniones de esta obra son de exclusiva responsabilidad de sus autores y no comprometen a la Universidad ni sus políticas institucionales.

Contenido

Prefacio	6
-----------------	----------

BIG DATA E INTELIGENCIA ARTIFICIAL APLICADA A LA SEGURIDAD

1. Definiendo Métricas para la Gestión Dinámica del riesgo, mediante la metodología MARISMA	8
Luis E. Sánchez, Antonio Santos-Olmo, Julio Moreno, Manuel Serrano, David G. Rosado, Eduardo Fernández-Medina	
2. Inteligencia Artificial aplicada al Análisis Forense Digital: Una revisión preliminar	19
J. J. Cano, J. D. Miranda, S. Pinzón	
3. MARISMA BiDa: Gestión y control del riesgo en Big Data. Caso de estudio	28
David G. Rosado, Julio Moreno, Luis E. Sánchez, Antonio Santos-Olmo, Manuel A. Serrano, Eduardo Fernández-Medina	

ANÁLISIS FORENSE APLICADO A SISTEMAS MULTIMEDIA

4. Detección de Manipulaciones Copy-Move en Ficheros Multimedia mediante la Transformada Discreta del Coseno	42
Esteban Alejandro Armas Vega, Ana Lucila Sandoval Orozco, Luis Javier García Villalba, Member, IEEE	
5. Detección de Post-Procesamiento en Vídeos Digitales de Dispositivos Móviles mediante el Análisis de la Estructura del Contenedor Multimedia	54
Carlos Quinto Huamán, Daniel Povedano Álvarez, Ana Lucila Sandoval Orozco, Luis Javier García Villalba, Member, IEEE	

6. **Técnica de Detección de Manipulación en Vídeos Digitales Basada en los Algoritmos de Compresión** 64
Edgar González Fernández, Ana Lucila Sandoval Orozco,
and Luis Javier García Villalba, Member, IEEE

TENDENCIAS EN SEGURIDAD

7. **Methodology to Securitize Smart Toys in Household Environments** 76
P. González, F. Paniagua, D. Suárez, J. J. Nombela
8. **Análisis de la Estructura del Contenedor de Vídeos Digitales de Dispositivos Móviles para identificación de la Fuente de Vídeos en Escenarios Abiertos** 85
Raquel Ramos López, Elena Almaraz Luengo, Ana Lucila Sandoval Orozco, Luis Javier García Villalba, Member IEEE

ANÁLISIS DE AMENAZAS Y VULNERABILIDADES

9. **Modelo SOCIA. Una reflexión conceptual y práctica desde la perspectiva del adversario** 97
Jeimy J. Cano M.
10. **Amenazas Cibernéticas a la Seguridad y Defensa Nacional. Reflexiones y perspectivas en Colombia** 105
M. E. Realpe, J. Cano

GESTIÓN DE LA SEGURIDAD Y ANÁLISIS DE RIESGOS

11. **Realizando una Revisión Sistemática de Metodologías ISRA orientadas a la Seguridad TIC. Periodo 2014-2019** 115
L. E. Sánchez, A. Santos-Olmo, V. Figueroa, D.G. Rosado,
E. Fernandez-Medina
12. **Análisis de la Privacidad y de la Seguridad en Contratos Inteligentes** 130
Antonio López Vivar, Ana Lucila Sandoval Orozco,
and Luis Javier García Villalba, Member IEEE

- 13. Revisión Sistemática de Análisis de Riesgos Asociativos y Jerárquicos. Periodo 2014 2019** **139**
A. Santos-Olmo, L. E. Sánchez, E. Álvarez,
D.G. Rosado, E. Fernandez-Medina

SEGURIDAD EN REDES

- 14. Análisis y evaluación de la Seguridad proporcionada por DNSSEC en Redes de Información IPv6 en un escenario de pruebas controlado** **149**
D. K Terán Arévalo, D. V. Fernández García, F. J Terán Cuarán
- 15. Vulnerability Analysis Oriented to Business Logic of Web Applications. A Case Study** **157**
William-Rogelio Marchand-Niño, Claudia Terrones Murga
- 16. Análisis y Monitorización por Agrupamiento de Contenido Multimedia en Redes SDN** **165**
Jesús Antonio Puente Fernández and Luis Javier
García Villalba, Member IEEE
- 17. Encryption with TLS Protocol version 1.2 and Web Sites Performance. A Case Study** **172**
William-Rogelio Marchan, Edgar Etson Rueda Liberato

TENDENCIAS DE LA EDUCACIÓN EN SEGURIDAD

- 18. Laboratory for Vulnerability Analysis and CIS Controls on Layer 2 Switches** **181**
William-Rogelio Marchand-Niño, José Martin Santillan Ruiz
- 19. Estrategias para la implementación y adopción de componentes curriculares de ciberseguridad en programas de grado y posgrado** **191**
F.A. Corredor, D. C. Franco, J. E. Martínez

Prefacio

El x Congreso Iberoamericano de Seguridad Informática - CIBSI 2020 se celebró del 22 al 24 de enero de 2020, en el antiguo Claustro en la Universidad Rosario, en la ciudad de Bogotá (Colombia). Fue organizado por el Departamento de Matemáticas Aplicadas y Ciencias de la Computación (MACC) de la Universidad del Rosario y CRIPTORED, con el apoyo de Internet Society – Colombia Chapter y Bizagi.

El objetivo principal del CIBSI fue promover y desarrollar el área de la Seguridad de la Información, generando un ámbito permanente de aprendizaje e intercambio de experiencias alrededor de esta disciplina, abordando temas y aspectos de actualidad que son la principal preocupación de quienes se desempeñan en el sector.

Se recibieron 37 artículos (full papers) y fue aceptado para presentarse en el congreso su 43%, es decir, 16 artículos. Todo ellos giraron en torno a las siguientes categorías:

- BIG-DATA e Inteligencia Artificial aplicada a la Seguridad
- Análisis Forense aplicado a Sistemas Multimedia
- Tendencias en Seguridad
- Análisis de Amenazas y Vulnerabilidades
- Gestión de la Seguridad y Análisis de Riesgos
- Seguridad en Redes
- Tendencias de la Educación en Seguridad.

Los editores

BIG DATA E INTELIGENCIA ARTIFICIAL APLICADA A LA SEGURIDAD

Definiendo Métricas para la Gestión Dinámica del riesgo, mediante la metodología MARISMA

Luis E. Sánchez, Antonio Santos-Olmo, Julio Moreno, Manuel Serrano, David G. Rosado, Eduardo Fernández-Medina

Resumen – En una sociedad en que los sistemas de Análisis y Gestión del riesgo son cada vez más necesarios para la correcta Gestión de la Seguridad de las compañías, surge la problemática de cómo hacer que estos sistemas pueden evolucionar y aprender de una forma natural, contando con la mayor información posible. Esta evolución del riesgo tiene que estar asociada con los eventos que ocurren dentro de la compañía, tanto los que pueden ser automatizados como los que deben gestionarse de forma manual, pero también debe poder asociarse con los riesgos externos que cambian con el tiempo. En este artículo presentamos algunas de las métricas que se han definido dentro de la metodología MARISMA (Methodology for the Analysis of Risks on Information System, using Meta-Pattern and Adaptability), con el objetivo de conseguir que los análisis de riesgos sean capaces de evolucionar de forma dinámica teniendo en cuenta elementos internos y externos, incluso los asociados con riesgos de terceras compañías.

Palabras clave — Cibersecurity, Information Systems Security Management, ISRA, Information Security Risk Analysis, ISMS, Risk Analysis.

I. INTRODUCCIÓN

Hoy en día, los elementos digitales o las infraestructuras (computadoras, redes, contenidos, etc.) son elementos cada vez más complejos y dependientes de las TI, que están en el centro de nuestras vidas y constituyen los pilares esenciales de nuestras infraestructuras de comunicación, económicas, sociales e institucionales [1]. La seguridad y la mitigación de amenazas dentro de esos sistemas se ha convertido implícitamente en una parte fundamental para el ciudadano (para preservar su privacidad), para la empresa (para proteger los activos y transacciones digitales) y para los estados (para proteger sus infraestructuras críticas y asegurar la continuidad del gobierno y servicios gubernamentales, etc.) [2, 3].

Esta seguridad es cada vez más importante en entornos como el sector salud, las PYMES, las Infraestructuras Críticas, las Industrias asociadas con la Industria 4.0, o en los propios ciudadanos con la aparición de conceptos como el IoT y los sistemas ciberfísicos [4, 5].

Para proteger estos sistemas recurrimos a la gestión de la seguridad, que según [6] puede definirse como un sistema de gestión usado para establecer y mantener un entorno seguro de la información. El objetivo principal de un SGSI (Sistema de Gestión de Seguridad de la Información) es afrontar la puesta en práctica y el mantenimiento de los procesos y procedimientos necesarios para manejar la seguridad de las tecnologías de la información. Estas acciones incluyen la identificación de las necesidades de seguridad de la información y la puesta en práctica de estrategias para satisfacer estas necesidades, medir los resultados y mejorar las estrategias de protección.

La definición de un SGSI es una tarea ardua y compleja que requiere un proceso previo de definición en la compañía donde se quiere establecer. Una de las fases más importantes para la implantación de un SGSI es la de Análisis y gestión de riesgos, que para algunos investigadores es una de las fases más críticas [7].

Todas las organizaciones que utilizan tecnologías de la información tienen problemas con la seguridad de su sistema de información. El primer paso en el proceso de protección de un sistema de información es la identificación y clasificación de los recursos o activos de información que necesitan protección, porque son vulnerables a las amenazas, y para realizar este paso necesitamos contar con sistemas de análisis y gestión de riesgos adecuados [8]. Diferentes investigadores destacan que la gestión de riesgos es un proceso esencial en cualquier modelo de gestión empresarial [9, 10], y que la información es un activo valioso que se espera que esté protegido.

Un análisis de riesgos es un proceso sistemático para estimar la magnitud de los riesgos a los que está expuesta una organización, para saber qué decisión tomar ante una posible eventualidad [11]. Para ello, se seleccionan e implementan salvaguardas para poder conocer, prevenir, impedir, reducir o controlar los riesgos identificados. Esto es lo que se entiende como gestión de riesgos.

De forma más técnica, el análisis de riesgos permite determinar cómo es, cuánto vale y cómo de protegidos se encuentran los activos. En coordinación con los objetivos, estrategia y política de la organización, las actividades de gestión de riesgos permiten elaborar un plan de seguridad que, implantado y operado, satisfaga los objetivos propuestos con el nivel de riesgo que acepta la dirección.

Actualmente se están realizando muchas investigaciones sobre análisis de riesgos, y muchas de ellas intentan comparar los métodos clásicos para ver cómo se podrían alinear [12-18]. Otros investigadores han realizado también algunos análisis comparativos de los principales estándares de riesgos con el

L. E. Sánchez, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, Luisenrique@sanchezcrespo.org

A. Santos-Olmo, Departamento I+D+i, Sicaman Nuevas Tecnologías y Marisma Shield, Tomelloso (Ciudad Real), España, Asolmo@sicaman-nt.com

Julio Moreno, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, Julio.moreno@uclm.es

Manuel A. Serrano, Grupo de Investigación Alarcos, Universidad de Castilla-la Mancha, Ciudad Real, España, Manuel.serrano@uclm.es

D.G. Rosado, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, David.Grosado@uclm.es

E. Fernandez-Medina, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, Eduardo.FdezMedina@uclm.es

objetivo de mejorar algunos de sus aspectos [19, 20].

Uno de los principales puntos que destacan las metodologías existentes, es la necesidad de poder contar con sistemas de riesgos que sean capaces de evolucionar y aprender de la experiencia. Así, Dehkhoda [21] analiza la posibilidad de unir métodos tradicionales de análisis de riesgos como IRAM2 con los análisis CBA (Cost-Benefit Analysis), ya que entiende que los unos no pueden funcionar sin los otros. Wangen [22] plantea la necesidad de contar con métricas que puedan ser valoradas en entornos profesionales. Según Pandey [23], cualquier activo de información, cuando está conectado al mundo exterior, es vulnerable a los ataques y estos ataques deben ser identificados lo antes posible. López [24] se centra en el Riesgo Dinámico y destaca que la aplicación de procesos de Análisis y Gestión de Riesgos en el ámbito de los Sistemas de Información es una práctica común que permite la planificación en un momento puntual de tiempo de las acciones preventivas frente al riesgo a corto, medio o largo plazo, pero con un considerable potencial actualmente desaprovechado para facilitar la toma de decisiones en tiempo real frente a eventos o incidentes de seguridad. Este trabajo hace un recorrido por las principales corrientes que buscan sacar partido a este potencial, englobadas principalmente bajo el concepto de Análisis de Riesgos Dinámico.

Otros investigadores no sólo destacan la necesidad de contar con sistemas dinámicos, sino también con sistemas que sean capaces de adaptarse a los entornos cambiantes. De esta forma, Saripalli [25] plantea la necesidad de modernizar los modelos de riesgos para que se puedan adaptar a entornos como el Cloud Computing [25], y Li y Sicari [26, 27] proponen la necesidad de avanzar en el análisis de riesgos para IoT, tanto desde el punto de vista de modernizar los modelos existentes como de la necesidad de añadir sistemas inteligentes de valoración del riesgo, tales como las redes neuronales.

Después de analizar estas propuestas, vemos que es necesario poder contar con sistemas y mecanismos que nos permitan generar sistemas de gestión de riesgos que tengan la capacidad de aprender y evolucionar de forma constante, mediante el análisis y la inclusión de todas las variables de su entorno. No sólo las variables internas, sino también las variables externas, y en este aspecto es importante incluir los ataques que sufren otras empresas de su entorno y tipología.

Para cumplir con todos estos aspectos, es necesario afrontar el desarrollo de una estructura de procesos que permita relacionar todos estos elementos y un conjunto de métricas que les dé soporte.

El resto del artículo se estructura de la siguiente forma: En la Sección 2 se analizarán los procesos principales que conforman la metodología MARISMA, y cómo estos elementos se relacionan entre sí para transmitir la información y el conocimiento generado. En la Sección 3, se presentarán el conjunto de métricas e indicadores iniciales que han sido desarrollados para cumplir con los objetivos de la investigación, explicando cada uno de los cinco bloques en que se dividen. Finalmente, en la última sección describiremos las principales conclusiones obtenidas.

II. METODOLOGÍA MARISMA

En este apartado explicaremos las bases sobre las que se sustenta la metodología de MARISMA, y para ello se mostrará una visión general de los procesos que conforman la metodología.

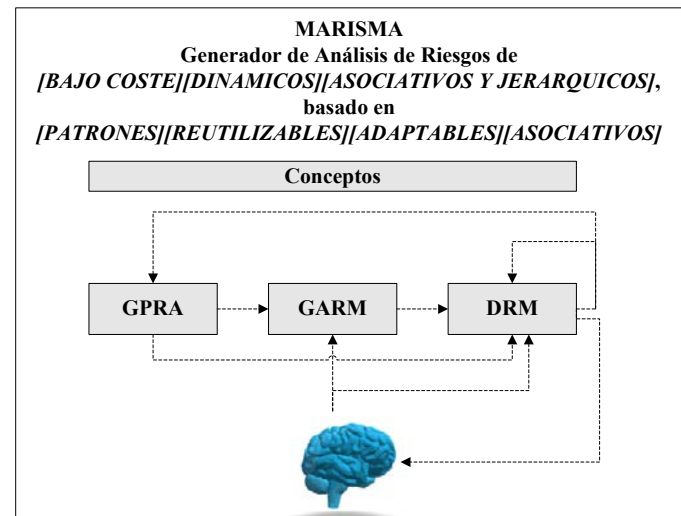


Figura - 1. Visión general de los procesos de la metodología

La metodología MARISMA está constituida por tres Procesos (ver Figura 1), que a su vez se dividen en actividades, tareas y sub-tareas:

- *Proceso 1 – Generación de Patrones para el Análisis de Riesgos (GPRA):* Se establece una estructura de relaciones entre los diferentes elementos involucrados en el análisis de riesgos y los controles necesarios para gestionar la seguridad. Estas relaciones se establecen mediante el conocimiento adquirido en las diferentes implantaciones, que es almacenado en una estructura denominada patrón para ser reutilizado con posterioridad, reduciendo los costes de generación de este proceso.
- *Proceso 2 – Generación del Análisis y Gestión del Riesgo (GARM):* Mediante la selección del patrón más adecuado y la identificación de un pequeño conjunto de los principales activos se obtiene un detallado mapa de la situación actual (análisis del riesgo) y un plan de recomendaciones de cómo mejorarlo (gestión del riesgo).
- *Proceso 3 – Mantenimiento Dinámico del Análisis de Riesgos (DRM):* Mediante la utilización de las matrices generadas, las cuáles interconectan los diferentes artefactos, el sistema irá recalculando el análisis de riesgos según se produzcan incidentes de seguridad, fallen las métricas definidas o los auditores detecten “no conformidades” en los controles.

En la Figura 2 se puede ver los tres procesos que componen la metodología MARISMA, y cómo intercambian información entre ellos para hacer evolucionar el sistema. A continuación explicaremos de forma resumida cada uno de los

elementos que la conforman, tomando como punto de partida la “Generación de un evento” en el tercero de los procesos, que es el elemento que provoca la “reacción evolutiva/adaptativa” en los demás:

- *Paso 1* – La generación de un evento en el proceso DRM provoca que la instancia de Análisis de Riesgos asociada al evento, generada mediante el proceso GARM, evolucione cambiando aspectos como el nivel de cobertura de un control, o la probabilidad de ocurrencia de una amenaza.
- *Paso 2* - La generación de un evento en el proceso DRM provoca cambios en el patrón asociado a la instancia que se creó mediante el proceso GPRA, permitiéndole reajustar las relaciones entre sus elementos.
- *Paso 3* - La generación de un evento en el proceso DRM provoca cambios en el patrón asociado a la instancia que se creó mediante el proceso GPRA, permitiéndole reajustar elementos asociados al riesgo externo temporal, ayudando a crear un escudo de seguridad global entre las compañías que usan ese patrón.
- *Paso 4* – Los cambios en el patrón provocados en el Paso 2 hacen que los patrones asociados por ley de herencia en las que se basa la metodología, también evolucionen, transmitiendo el conocimiento adquirido.
- *Paso 5* – Los cambios/evolución de todos los patrones, se transmiten a todas las instancias para ayudarles a mejorar y evolucionar.

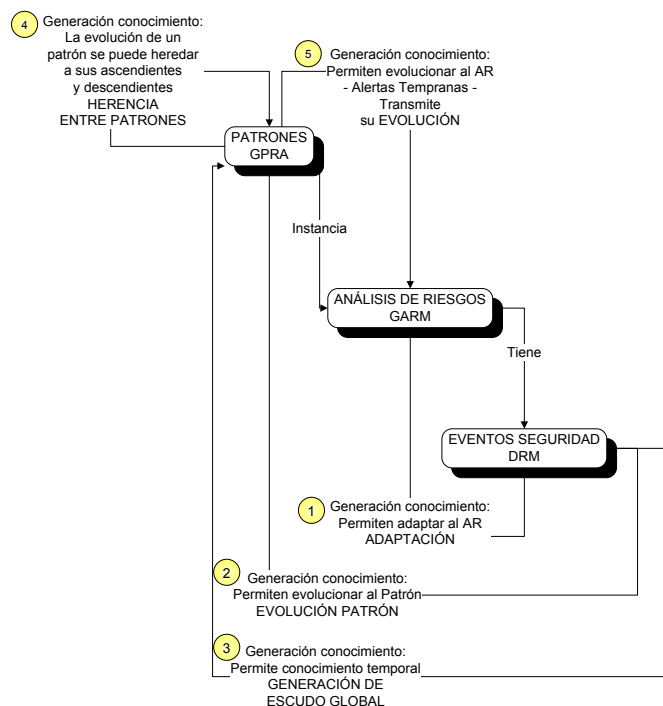


Figura - 2. Flujo de Generación de Conocimiento en MARISMA

Este ciclo permite, mediante la utilización de estos tres procesos, una constante evolución de las instancias

(individuos) y de los propios patrones (especie), haciendo que el sistema siempre evolucione con los cambios.

Estos tres procesos se dividirán en actividades y tareas, generando artefactos y relaciones entre ellas. La construcción de estos tres procesos, busca solucionar de forma total/parcial los problemas identificados durante la investigación.

III. MÉTRICAS PARA EL AR EN MARISMA

En este apartado se expone el conjunto de indicadores, reglas de negocio y métricas que han sido desarrolladas para poder dotar de dinamismo a la metodología MARISMA,

El objetivo último perseguido en esta fase es ser capaces de localizar y desarrollar indicadores y métricas que nos permitan calcular de forma semi-automática los valores asociados al nivel de riesgo, reduciendo el nivel de incertidumbre en la elaboración del análisis de riesgos.

Dentro del conjunto de métricas iniciales asociadas con la metodología MARISMA podemos distinguir cinco bloques:

- *Métricas asociadas al Patrón:* Son el conjunto de métricas que permitirán establecer relaciones de dependencia entre de los diferentes elementos del patrón (Ej: las dependencias entre los controles, las amenazas y los activos).
- *Métricas asociadas al Análisis de Riesgos:* Son el conjunto de métricas que utilizaremos para reajustar los valores asociados al Análisis de Riesgo y para realizar el cálculo del riesgo.
- *Métricas asociadas al Escudo de Seguridad Global (Global Security Shield – GSS):* Son el conjunto de métricas que nos permitirán crear una visión global del riesgo de cada elemento en un instante de tiempo dado, de forma asociativa entre todas las instancias relacionadas.
- *Métricas asociadas con la Sondas:* Son métricas asociadas con sondas (herramientas externas) que permiten el seguimiento automatizado de vulnerabilidades asociadas con controles.
- *Métricas asociadas a la valoración de los Activos:* Son métricas que nos permitirán calcular, de la forma más objetiva posible, el valor que tiene un activo de información para una compañía.

A continuación analizaremos las métricas de cada uno de estos bloques.

A. Métricas asociadas al Patrón.

Los patrones en Marisma, se basan en tres elementos principales (Controles, Activos y Amenazas) y existen relaciones entre ellos, tanto consigo mismos (Amenaza x Amenaza), como entre elementos de diferentes naturaleza (Control x Amenaza). Las métricas asociadas con el patrón tienen por objetivo determinar los grados de vinculación entre los elementos del patrón. Así por ejemplo, las dependencias entre Controles x Controles reflejaran una estadística de las dependencias asociadas a los eventos ocurridos por las instancias relacionadas con ese patrón.

Se definen inicialmente tres métricas para este tipo de relaciones:

- *Asociadas con los Controles:*
 - *Porcentaje de Inter-dependencias entre controles - Percentage of Inter-dependencies Between Controls (PIBC):* Define la relación que existe entre dos controles.
 - $PIBC = [n^{\circ} \text{ eventos en que aparecen ambos controles}] * 100 / [n^{\circ} \text{ eventos}]$
- *Asociadas con los Tipos de Activos:*
 - *Porcentaje de Inter-dependencias entre tipos de activos - Percentage of Inter-dependencies Between Types of Assets (PIBTA):* Define la relación que existe entre dos tipos de activos.
 - $PIBTA = [n^{\circ} \text{ eventos en que aparecen ambos tipos de activos}] * 100 / [n^{\circ} \text{ eventos}]$
- *Asociadas con las Amenazas:*
 - *Porcentaje de Inter-dependencias entre amenazas - Percentage of inter-dependencies between threats (PIBT):* Define la relación que existe entre dos amenazas.
 - $PIBT = [n^{\circ} \text{ eventos que en aparecen ambas amenazas}] * 100 / [n^{\circ} \text{ eventos}]$

Las dependencias entre matrices permitirán conocer mucho mejor la dependencia entre este tipo de elementos y ser la base para el desarrollo de un sistema experto que recomiende elementos complementarios a un usuario al rellenar un evento (Ej: Si el nivel de dependencia entre dos tipos de controles es del 75%, y en un evento añadimos uno de ellos, el sistema podrá indicar al usuario si está seguro de que el otro control no ha estado involucrado también en el evento de seguridad).

El segundo conjunto de métricas asociadas con esta parte estará orientada las relaciones inter-objetos, es decir entre objetos diferentes:

- *Métricas asociadas con las Amenazas y Controles:*
 - *Porcentaje de Inter-dependencias entre una amenaza y un control - Percentage of Inter-Dependencies Between a Threat and a Control (PIDBTC):* Define la relación que existe entre una amenaza y un control. Esta métrica es no modificable y sólo se mostrará sus resultados a nivel informativo.
 - $PIDBTC = [N^{\circ} \text{ de Amenazas en un evento que han activado ese Control}] * 100 / [n^{\circ} \text{ de veces que esa amenaza ha ocurrido en un evento}]$
 - *Porcentaje de Inter-dependencias entre una amenaza y un control por evento - Percentage of Inter-dependencies between a threat and an event control (PIBTEC):*

- $PIBTEC = \text{Al ocurrir una amenaza en un evento, para cada control involucrado:}$

⇒ Si existe la relación => Valor Actual + 1% (hasta un max de 100%).

⇒ Si no existe la relación => Valor Actual = 1%.

⇒ Para aquellos controles que están relacionados con la amenaza en la matriz pero no en el evento. Valor Actual – 1%.

- *Métricas asociadas con los Tipos de Activos, Amenazas y Dimensiones:*

- *Valor del Porcentaje de Degradación de la Dimensión de un Activo - Value of the Degradation Percentage of the Dimension of an Asset (VDPDA).*

- $VDPDA = \text{Valor medio del rango del evento.}$

B. Métricas asociadas al Análisis de Riesgos.

Una instancia de un análisis de riesgos debe tener dos objetivos principales:

- Capacidad de calcular el riesgo.
- Capacidad de adaptación del riesgo y sus elementos ante cambios.

Para calcular el riesgo, existen diferentes fórmulas que se pueden utilizar, aunque inicialmente la metodología MARISMA ha configurado la siguiente:

- $\text{Nivel Riesgo Actual} = ([\text{Valor Activo}] \times [\text{Max (Impacto Dimensiones)}]) * [\text{Probabilidad de ocurrencia}] * [\text{Vulnerabilidad}]$, calculándose la vulnerabilidad a partir de la ausencia del nivel de cobertura de los controles. El nivel de riesgo se mueve en un rango de [1-500].

El Nivel de Riesgo Actual se transformará a una Escala de Riesgo de 10 niveles, en base a una escala logarítmica, que se calcula aplicando la fórmula indicada en la Ecuación 1.

$$\begin{aligned}
 y &= a^x \\
 500 &= a^{10} \\
 \log 500 &= 10 \log a \\
 \log a &= \frac{\log 500}{10} = 0,26989700 \\
 a &= 10^{0,26989} = 1,8616455 \\
 y &= a^x \rightarrow x = \frac{\log y}{\log 1,86164}
 \end{aligned}$$

Ecuación 1. Cálculo de niveles para riesgo.

Los resultados de aplicar la Ecuación 1 sobre una escala de riesgo máximo de 500 se puede ver de forma numérica y gráfica en la Figura 3.

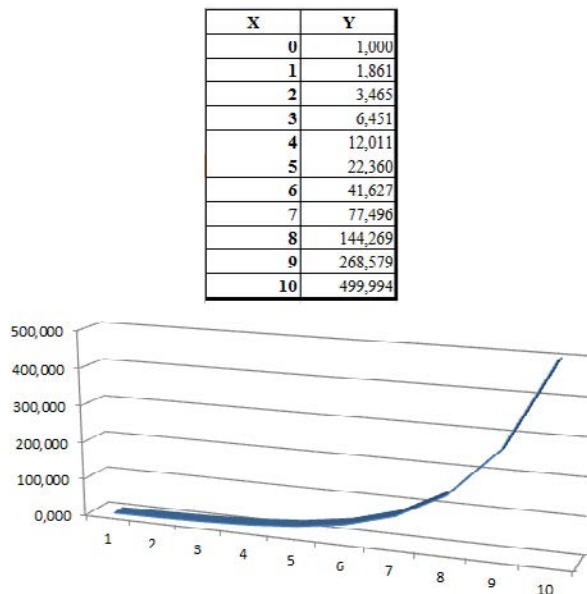


Figura - 3. Escala logarítmica para el cálculo de niveles de riesgo.

Este es sólo un ejemplo de métrica que nos puede permitir calcular el riesgo. Existen otras métricas que podríamos llegar a aplicar, por lo que se ha buscado que la metodología MARISMA extraiga los valores de los niveles de controles, vulnerabilidades, valoración de activos y amenazas, con el objetivo de poder hacer extensivas sus métricas, de forma que dé cabida a las métricas recomendadas por otros estándares, como:

- CRAMM (Central computer and Telecommunication Agency Risk Analysis and Management Method) [28, 29]:
 - $CRAMM = [\text{Riesgo (Amenaza, Activo)} = \text{Probabilidad (Amenaza)} * \text{Vulnerabilidad (Amenaza, Activo)} * \text{Impacto (Amenaza, Activo)}]$
- MEHARI (Méthode Harmonisée d'Analyse de Risques—Harmonised Risk Analysis Method) [30, 31]:
 - $MEHARI = [\text{Riesgo (Amenaza, Activo)} = \text{Probabilidad (Amenaza)} * \text{Vulnerabilidad (Amenaza, Activo)} * \text{Impacto (Amenaza, Activo)}]$
- OCTAVE (The Operationally Critical Threat, Asset, and Vulnerability Evaluation) [32]:
 - $OCTAVE = [\text{Riesgo (Amenaza, ActivoCritico)} = \text{Impacto (Amenaza, ActivoCritico)} * \text{Vulnerabilidad (ActivoCritico)}]$
- NIST SP 800-30 (Risk Management Guide for Information Technology Systems) [33]:

- $NIST\ SP\ 800-30 = [\text{Riesgo (Amenaza, Activo)} = \text{Probabilidad (Amenaza)} * \text{Vulnerabilidad (Amenaza, Activo)} * \text{Impacto (Amenaza, Activo)}]$

- CORAS (Construct a platform for Risk Analysis of Security Critical Systems) [34]:

- $CORAS = [\text{Riesgo (Amenaza, ActivoCritico)} = \text{Impacto (Amenaza, ActivoCritico)} * \text{Vulnerabilidad (ActivoCritico)}]$

En otros casos, estamos investigando cómo añadir algunas variables diferenciadas de las metodologías, como las “necesidades” en el caso de EBIOS, las “ExpectativaPérdidaAnual” de ISAMM, o el concepto de “Incidente” propuesto por IT-Grundschtz y por AS/NZS 4360, que podría encajar en el concepto de “Evento” propuesto por MARISMA:

- EBIOS (Expression of Needs and Identification of Security Objectives) [35]:
 - $EBIOS = [\text{Riesgo (amenaza, activo, necesidades)} = \text{Impacto (amenaza, necesidades)} * \text{Vulnerabilidad (amenaza, activo)}]$
- ISAMM (Information Security Assessment and Monitoring Method) [36]:
 - $ISAMM = [\text{Riesgo (Amenaza, Activo)} = \text{ExpectativaPérdidaAnual (Amenaza, Activo)} = \text{Probabilidad (Amenaza, Activo)} * \text{PromedioPérdida (Amenaza, Activo)}]$
- IT-Grundschtz: (IT Baseline Protection Manual) / BSI Standard 100-[2-3]) [37]:
 - $IT-Grundschtz = [\text{Riesgo (Incidente, Activo)} = \text{Probabilidad (Incidente)} * \text{Consecuencias (Incidente, Activo)}]$
- AS/NZS 4360 [38] & ISO31000:2018 [39]:
 - $AS/NZS\ 4360 = [\text{Riesgo (Incidente, Activo)} = \text{Probabilidad (Incidente)} * \text{Consecuencias (Incidente, Activo)}]$

Por otro lado, una vez que tenemos calculado el nivel de riesgo es importante tener la capacidad de aprender de los eventos, y desarrollar métricas que nos permitan re-alinear el valor subjetivo de los elementos iniciales para ajustarlos a la realidad.

Las métricas que forman parte de este apartado son:

- Métricas asociadas al Valor de Amenaza – Probabilidad de Ocurrencia:
 - *Valor Actual Recomendado para la Probabilidad de Ocurrencia de la Amenaza - Recommended Actual Value for the Probability of Occurrence of the Threat (RAVPOT):* Mediante un evento de seguridad podemos determinar si el valor actual de la

probabilidad de ocurrencia de una amenaza es correcto, o debería re-alinearse.

Para ello lo primero que haremos será considerar establecer rangos de valor para esa variable (Ej: en el caso de MARISMA será “1 día” la unidad base). De esta forma, si consideramos que la probabilidad de ocurrencia de una amenaza es “Alta si ocurre varias veces al mes” y “Muy Alta si ocurre varias veces al día”, e inicialmente la tenemos configurada como “Alta”, y tenemos notificados varios eventos relacionados con esa amenaza en un día, el sistema podrá recomendar el cambio de valor.

- *RAVPOT = Ante un Evento => Recalcular la probabilidad de ocurrencia en jornadas y compararla con los niveles definidos en los rangos*
 - ⇒ *Si el Rango es diferente al actual sugerir un cambio de nivel.*
 - ⇒ *Sino finalizar*

El conjunto de métricas presentado en este apartado es un conjunto inicial que se irá complementando con el tiempo según evolucione la metodología.

C. Métricas asociadas al GSS.

El Escudo de Seguridad Global (GSS) es un concepto que pretende unificar el poder de agregación de las redes colaborativas para informar a las instancias del riesgo temporal en cada momento sobre diferentes elementos.

Dicho de otra forma, pretendemos que un evento de seguridad que impacte sobre una compañía, pueda ser utilizado para que otras compañías que utilizan ese mismo patrón aumenten sus niveles de protección, generando un escudo de seguridad entre todas las compañías.

Para ello, inicialmente se tienen que definir una serie de dimensiones temporales y geográficas sobre las que calcularemos los datos agregados de las diferentes instancias. Inicialmente se han determinado los siguientes valores:

- *Rangos Temporales:* MuyCorto = Diario, Corto = Semanal, Medio = Mensual, Largo = Trimestral, MuyLargo=Anual.
- *Rangos Geograficos:* Continente, País, Región.

También se han definido un conjunto de métricas que estarán asociadas con cada uno de los principales elementos del Análisis de Riesgos:

- *Control:*
 - *Media del Riesgo Temporal y Externo asociado al Control - Average of the*

Temporary and External Risk associated to the Control (ATERAC).

- *ATERAC = N° de Eventos que ha incluido el control por ud Tiempo (Rangos) x Gravedad del Evento (leve o grave).*
- *Nivel Medio de Implementación del Control - Average Level of Implementation of the Control (ALIC).*
 - *ALIC= Nivel de cobertura media de los controles por ud Tiempo.*
- *Activos:*
 - *Media del Riesgo Temporal y Externo asociado al Tipo de Activo - Average of Temporary and External Risk associated with the Type of Asset (ATERATA).*
 - *ATERATA = N° de Eventos que han incluido el Tipo de Activo por ud Tiempo (Rangos) x Gravedad del Evento (leve o grave).*
 - *Valor medio de los Tipos de Activos - Average value of the Types of Assets (AVTA).*
 - *AVTA = Valor medio de los T.Activo por ud Tiempo.*
- *Amenazas:*
 - *Media del Riesgo Temporal y Externo asociado a la Amenaza - Average of Temporary and External Risk associated with the Threat (ATERAT)*
 - *ATERAT = N° de Eventos que han incluido la Amenaza por ud Tiempo (Rangos) x Gravedad del Evento (leve o grave).*
 - *Media de la Probabilidad de ocurrencia de la Amenaza - Average Probability of occurrence of the threat (APOT).*
 - *APOT = Valor medio de la Probabilidad de Ocurrencia de la Amenaza por ud Tiempo (Rangos).*
 - *Media de la Degradación del Valor Asociado a la Amenaza - Average Degradation of Value Associated with the Threat (ADVAT)*
 - *ADVAT = Valor medio del % de Degradación por ud Tiempo (Rangos).*

Cada evento de seguridad provocará una serie de cambios dinámicos tanto en su propio Análisis de Riesgos, como asociados al patrón del que fue generado. Este proceso se inicia, con la actualización de los valores asociados con su patrón, y éste a la vez provocará en cascada la actualización de todas sus dependencias. El valor cambiará la visión de cada

una de las tablas y se transmitirá a todas las instancias involucradas.

D. Métricas asociadas a Sondas.

Finalmente, se ha considera que es de gran interés poder asociar los patrones, y en concreto elementos como los “Controles”, con herramientas externas que complementen a los “Eventos Manuales” (Incidentes de seguridad reflejados por los usuarios del Sistema de Información), mediante revisiones automatizadas de vulnerabilidades en activos.

Para ello se está trabajando en métricas que permitan establecer esas relaciones de forma automatizada, permitiendo el alineamiento entre estándares (Ej: ISO27001) y los sistemas clásicos de análisis de vulnerabilidades (Ej: Vulnerabilidades BBDD CVE).

Por ejemplo, imaginemos que deseamos generar la capacidad de que un activo de tipo “Software” (SW, según el catálogo de MAGERIT) incluido dentro del alcance de nuestro análisis de riesgos, para un patrón basado en la ISO27001, adquiera la capacidad de monitorizar periódicamente si dicho activo está seguro o no con respecto a un conjunto de nuevas vulnerabilidades.

Actualmente las empresas, incluso las grandes, resuelven este problema de una forma “artesanal”. De esta forma, suelen indicar manualmente a un tercero (Ej: CNPIC) las tecnologías utilizadas en un software de valor para ellos, y este tercero de confianza les reporta mensualmente informe en el que indica si han aparecido nuevas vulnerabilidades relacionadas con dicho software. Una vez revisada, se debe buscar el nivel de riesgo de cada vulnerabilidad y determinar si les puede impactar o no. Dicho proceso está asimismo totalmente aislado del Análisis de riesgos, por lo que aunque existe un cambio en el riesgo asociado a dicho activo, no se traslada al informe de riesgos. De igual forma, el riesgo estará activo durante una ventana de tiempo que puede llegar incluso al mes de duración, algo que no es aceptable en la mayor parte de los casos.

El planteamiento de la metodología MARISMA parte de la creación de un “motor de sondas” que permita conectar los elementos base del patrón, con los elementos externos asociados a las sondas. Por ejemplo, en la Figura 4 podemos ver cómo:

- Por un lado, tomando de referencia el Patrón para Gestión de la Seguridad bajo la normativa de la ISO27001 podemos relacionar un Tipo de Activo del catálogo de MAGERIT (que sería el [SW] – Software) con la familia de amenazas [A] – Ataques Intencionados, y se podría considerar que afectaría al nivel de cumplimiento del control “A.12.2.1 – Controles contra el software malicioso”, disminuyendo el nivel de cobertura de este control ante la aparición de nuevas vulnerabilidades en activos software.
- Por otro lado, la sonda se nutre de la BBDD de vulnerabilidades CVE (Common Vulnerabilities and

Exposures) y de la métrica CVSS (Common Vulnerability Score System) que determina el nivel de gravedad de la vulnerabilidad, de tal forma que se penalizará más o menos el nivel de cobertura del control dependiendo del valor de CVSS. Finalmente, integraremos un extractor de tecnología como BuiltWith que permitirá que, dada una página web o un software en la nube, automáticamente extraerá las tecnologías utilizadas para su desarrollo y las versiones de éstas.

Estas herramientas no siempre nos permitirán automatizar el 100% del análisis de vulnerabilidades, pero reducen de forma considerable el trabajo necesario para controlar la aparición de nuevas vulnerabilidades vinculadas a los activos software, y permiten de forma automática cambiar el nivel de riesgo asociado a los activos de información, permitiendo tomar decisiones sobre cómo afrontar este nuevo riesgo de una forma ágil.

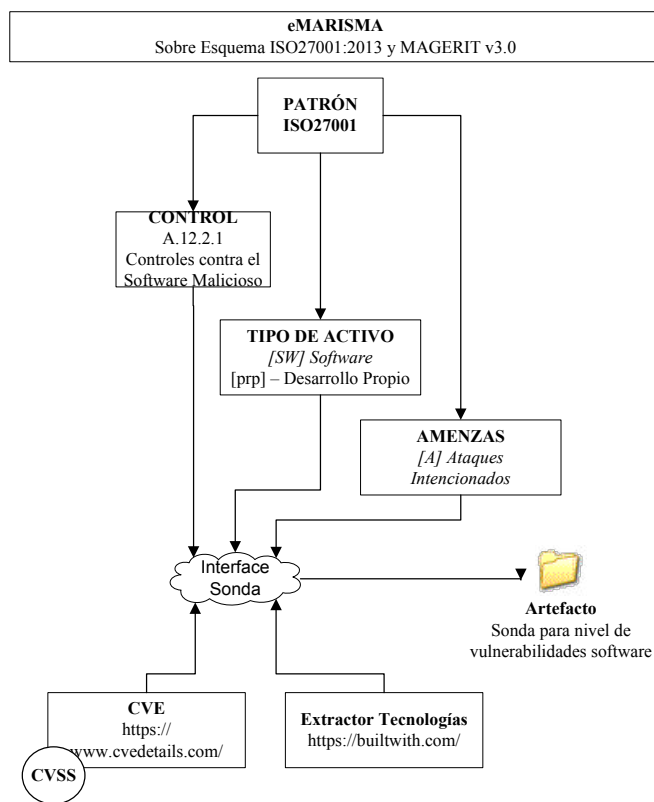


Figura - 4. Ejemplo de sonda

E. Métricas asociadas a la Valoración de Activos.

Uno de los grandes problemas cuando intentamos valorar el riesgo asociado a los activos de información de una compañía es ser capaces de encontrar mecanismos para calcular el valor monetario de estos de la forma más objetiva y precisa posible, dado que este valor es de vital importancia para estimar las pérdidas potenciales debidas a impactos de

seguridad sobre dichos activos.

Dentro de los mecanismos de valoración de activos podemos destacar el método propuesto por Thomas A. Stewart [40] que, en su artículo "Trying to grasp the intangible", propone una metodología ideada por la empresa norteamericana NCI y orientada a medir el valor de los "activos intangibles" de una empresa, que incluirían no sólo la parte de activos TIC (software, claves criptográficas, etc.), sino también otros activos de valor (know-how, valor de la marca, patentes, conocimiento de los mercados, etc.). De esta forma, el método ofrece una orientación sobre cómo medir el valor de los activos en información de una organización.

El método propuesto está formado por siete partes:

1. Paso 1 - Se toma el promedio durante los tres últimos años del beneficio antes de impuestos.
2. A partir del balance de la empresa se toma el promedio, en el mismo período, del valor de los activos tangibles.
3. Se dividen los beneficios del punto 1 por el valor de los activos tangibles del punto 2, y se pasa a %, obteniéndose así el rendimiento sobre activos (ROA).
4. Para el mismo período, se calcula el promedio del ROA para las empresas del sector de la compañía considerada. Si el valor del ROA de una empresa es inferior al ROA promedio del sector, el método deja de aplicarse.
5. Se calcula lo que debería haber ganado la empresa (beneficios antes de impuestos) si hubiera tenido un ROA como el promedio del sector. Esta cantidad es el "exceso" de beneficio que ha tenido respecto a las empresas del sector, y que no se debe (si se ha seguido el razonamiento del método) a los activos tangibles.
6. Se deducen los impuestos, para determinar el "exceso" de beneficios después de impuestos.
7. Finalmente, se calcula cuánto debería tener la empresa en capital para que generara esta cantidad como réditos. Eso nos permitirá obtener un valor global de los activos intangibles de la compañía.

Aunque el método tiene sus ventajas presenta algunas desventajas, como que sólo permite obtener una estimación de un valor global de los activos intangibles de la empresa, pero no su valor individualizado, y por otro lado es difícil de aplicar en PYMES. Aun así, consideramos que puede ser de interés su aplicación en ciertos casos.

Por otro lado, es importante tener en cuenta qué tipos de activos vamos a tener dentro de la compañía, y cómo se podría llegar a valorar cada uno de ellos. Para realizar una primera aproximación a la valoración de los activos utilizaremos el catálogo de activos de MAGERIT [11]:

- *[D] Datos / Información:* Los datos son el corazón que permite a una organización prestar sus servicios.

Este es uno de los tipos de activos más difíciles de cuantificar dentro de una compañía, y requiere de un estudio profundo de investigación que queda fuera del alcance de este artículo, por lo que nos limitaremos a plantar algunas fórmulas sencillas de valoración:

- *Formula1:* La utilizaremos para valorar los ficheros y los datos de BBDD de valor en la compañía, entendiendo que estos hayan sido obtenidos de forma de legal y, por lo tanto, sean declarables por la misma. El cálculo se puede realizar teniendo en cuenta el % de beneficio asociado a esa BBDD (estimado por el departamento contable de la compañía), el número de registros de esa BBDD y aplicando un factor de amortización del 10% según la antigüedad de los datos, ya que cuanto más antiguos sean los datos mayor riesgo existe de que hayan perdido valor de confidencialidad o de generación de conocimiento.

$$\begin{aligned} \text{Valor Activo} &= [\% \text{ BeneficioAsociadoBBDD}] * \\ &[\text{DepreciaciónAsociadaPorRegistro}] \end{aligned}$$

- *Formula2:* Asociada a código fuente desarrollado internamente. Trataremos de dar un precio a un software desarrollado internamente asociándolo al número de horas que se han invertido, por el coste hora, teniendo en cuenta que un 20% del esfuerzo anual se amortiza por depreciación tecnológica. Igualmente, debemos ejercer un factor correctivo según el porcentaje del activo asociado a los beneficios de la compañía (Ej: Un software que se ha desarrollado en tres años con un coste hora de 25€, y un inversión en horas cada año de [100,50,80], tendrá un valor de 3.680 € el tercer año, y si está asociado al 20% del beneficio de la compañía tendrá un valor final de 4.416 €).

$$\begin{aligned} \text{Valor Activo} &= [\text{N}^\circ \text{ Horas Invertidas} \\ &\text{Desarrollo}] * [\text{Coste Hora}] * [\% \\ &\text{DepreciaciónTecnologica}] * [\% \\ &\text{AsociadoBeneficios}] \end{aligned}$$

Existen otras fórmulas que podemos implementar para este tipo de activos (Ej: $[\text{Valor Sector}] * [\text{N}^\circ \text{ Registros Activo}]$, $[\text{Valor Global de Intangibles de la compañía}] - [\text{Valor del Resto de Activos Intangibles menos datos}]$), que podrían obtenerse a partir de preguntas del tipo "¿Sector al que

pertenece la compañía?”, “¿Número de registros del activo? (Ej: proveedores, clientes, ...)”, etc.

Otras características que se podrían valorar para realizar estos cálculos son: i) Calidad de los datos; ii) Si se tienen registrados y documentados legalmente (LOPD); iii) Número de consultas por mes realizadas; iv) % de la facturación de la compañía asociado con ellos; v) Número de empleados que lo utilizan; vi) Número de horas que se usan esos ficheros al año; vii) Antigüedad media de los datos; viii) Naturaleza del dato (Público, privado, confidencial).

- *[K] Claves criptográficas:* La criptografía se emplea para proteger el secreto o autenticar a las partes. Podemos considerar que el valor de una clave está vinculada al valor de la información que protege, o de los beneficios asociados a la misma.

- *Formula1:* Valor asociado a la información que protege.

- *[S] Servicios:* Función que satisface una necesidad de los usuarios (del servicio). El valor de un servicio está directamente asociado al % Beneficios que genera en la compañía, y se puede regir por principios contables.

- *Formula1:* % Beneficios asociados al servicio.

- *[SW] Software - Aplicaciones informáticas:* El valor de un software desarrollado por terceros, o con licencia comercial, se puede regir por principios contables, y en el caso de un código desarrollado internamente por la formula desarrollada en el apartado “[D] – Datos”.

- *[HW] Equipamiento informático (hardware):* Dícese de los medios materiales, físicos, destinados a soportar directa o indirectamente los servicios que presta la organización, siendo pues depositarios temporales o permanentes de los datos, soporte de ejecución de las aplicaciones informáticas o responsables del procesado o la transmisión de datos.

- *Formula1:* Podemos calcular su valor por principios contables: Coste * Amortización.

- *[COM] Redes de comunicaciones:* Incluyen tanto instalaciones dedicadas como servicios de comunicaciones contratados a terceros; pero siempre centrándose en que son medios de transporte que llevan datos de un sitio a otro.

- *Formula1:* Podemos calcular su valor por principios contables: Coste * Amortización.

- *[L] Instalaciones:* En este epígrafe entran los lugares donde se hospedan los sistemas de información y comunicaciones.

- *Formula1:* Podemos calcular su valor por principios contables: Coste * Amortización.

- *[P] Personal:* En este epígrafe aparecen las personas relacionadas con los sistemas de información. En este caso el coste se puede calcular teniendo en cuenta el KnowHow de esa persona asociado al beneficio de la compañía, y con un máximo de antigüedad de 10 años, dado que los conocimientos superiores a 10 años rara vez son útiles en la tecnología.

- *Formula1:* En este caso, el coste se puede calcular teniendo en cuenta el KnowHow de esa persona asociado al beneficio de la compañía, y con un máximo de antigüedad de 10 años, y multiplicado por el salario medio del perfil que tendría que ser reemplazado.

$$\begin{array}{lcl} \text{Valor} & \text{Activo} & = \\ \text{[SalarioMedioPerfil]} & * & \text{[%} \\ \text{BeneficioAsociadoRRHH]} & & * \\ \text{[FactorCorrectorNumeroAños]} & & \end{array}$$

Este conjunto de métricas debe ser tomado como una base que ira siendo cumplimentada en posteriores investigaciones, según se vaya disponiendo de más casos de uso para su validación.

IV. CONCLUSIONES.

En este artículo se ha analizado el conjunto de métricas e indicadores que se ha desarrollado como parte de la metodología MARISMA para dotar de capacidad de dinamismo y aprendizaje a los análisis de riesgos.

Se han introducido conceptos como el riesgo asociativo entre compañías, que permite que los eventos de seguridad de una compañía puedan ser utilizados por otras compañías para aumentar su nivel de protección.

El conjunto de métricas presentado en el artículo ha sido implementado dentro de la herramienta eMARISMA que da soporte a la metodología MARISMA, y está siendo validado en casos reales dentro del entorno de la “Gestión de la Seguridad” utilizando estándares como la ISO27001 y Magerit, así como la Ley 8/2011 de Infraestructuras Críticas.

Como líneas futuras se está complementando el conjunto de métricas, añadiendo nuevas métricas que puedan generar inteligencia en el sistema, y se está desarrollando una versión de la herramienta que permitirá de forma gratuita que los usuarios puedan utilizar las métricas para valorarlas y generar una base de conocimiento común, que permita aumentar el nivel de protección de cada una de las compañías de forma individual, pero utilizando el conocimiento colectivo de todas.

Podemos concluir, por lo tanto, que ahora mismo es muy importante poder disponer de sistemas de análisis de riesgos dinámicos y que permitan compartir riesgos y colaborar entre diferentes empresas, y que las métricas presentadas en este artículo como parte de la metodología MARISMA son una primera aproximación para conseguir ese objetivo.

AGRADECIMIENTOS

Esta investigación ha sido co-financiada por los proyectos *GENESIS - Security Government of Big Data and Cyber Physics Systems ((SBPLY/17/180501/000202)* financiado por el “JCCM- Consejería de Educación, Cultura y Deportes, y Fondos FEDER”, del proyecto *ECLIPSE – Enhancing Data Quality and Security for Improving Business Processes and Strategic Decisions in Cyber Physical Systems (RTI2018-094283-B-C31)* financiado por la “Ministerio Economía, Industria y Competitividad y fondos FEDER”, y ha contado con el apoyo de las empresas Marisma Shield S.L (www.emarisma.com) y Sicaman Nuevas Tecnologías (www.sicaman-nt.com).

REFERENCIAS

- [1] Huerta, M., et al. *Design of a building security system in a university campus using RFID technology*. in *2017 IEEE 37th Central America and Panama Convention (CONCAPAN XXXVII)*. 2017. IEEE.
- [2] Le Grand, G. and E. Adar. *White cyber knight—a Risk Assessment tool for network resilience evaluation*. in *the proceedings of the International Workshop on Complex Network and Infrastructure Protection (CNIP'06), Rome*. 2006.
- [3] Vivas, T., A. Zambrano, and M. Huerta. *Mechanisms of security based on digital certificates applied in a telemedicine network*. in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2008. IEEE.
- [4] Huerta, M., et al. *Implementation of a open source security software platform in a telemedicine network*. in *Proceedings of the 9th WSEAS international conference on Advances in e-activities, information security and privacy*. 2010. World Scientific and Engineering Academy and Society (WSEAS).
- [5] Pirrone, J. and M. Huerta. *Security Mechanism for Medical Record Exchange Using Hippocratic Protocol*. in *World Congress on Medical Physics and Biomedical Engineering 2018*. 2019. Springer.
- [6] Eloff, J. and M. Eloff, *Information Security Management - A New Paradigm*. Annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology SAICSIT'03, 2003: p. 130-136.
- [7] Steve, E., *An Introduction to information systems risk management*. SANS Institute InfoSec Reading Room. Available on-line at http://www.sans.org/reading_room/whitepapers/auditing/introduction-information-system-risk-management_1204. Accessed March, 2006. **16**: p. 2011.
- [8] Bača, M. and F. Varaždin, *The risk assessment of information system security*. Fakultet organizacije i informatike, Sveučilište u Zagrebu. < dostupno na http://cuc.carnet.hr/cuc2004/program/radovi/a5_baca/a5_full.pdf>, [očitano 07.10. 2010], 2004.
- [9] Restrepo, L.O. and F.J.V. Duque, *Gestión de riesgos en eTOM. Un análisis comparativo con los estándares de riesgo corporativo*. Revista Logos, Ciencia & Tecnología, 2017. **9**(1): p. 85-99.
- [10] Ortiz Restrepo, L., V. Duque, and F. Javier, *Gestión de riesgos en eTOM. Un análisis comparativo con los estándares de riesgo corporativo*. Revista Logos Ciencia & Tecnología, 2017. **9**(1): p. 85-99.
- [11] Magerit V3, *Methodology for Information Systems Risk Analysis and Management*., in *Ministerio de Hacienda y Administraciones Públicas* 2012: Spain.
- [12] Wangen, G., *Information security risk assessment: a method comparison*. Computer, 2017. **50**(4): p. 52-61.
- [13] Fakrane, C. and B. Regragui. *Interactions and Comparison of It Risk Analysis Methods*. in *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech)*. 2018. IEEE.
- [14] Jeannot, F., *Méthodologies d'évaluation et gestion de risques en sécurité*. Montréal, Canada, Mai 2018, R518, v1.0. 2018.
- [15] Benavides Sepúlveda, A.M. and C.A. Blandón Jaramillo, *Modelo de sistema de gestión de seguridad de la información basado en la norma NTC ISO/IEC 27001 para instituciones públicas de educación básica de la comuna Universidad de la ciudad de Pereira Alejandra*. 2017.
- [16] Bommman, W.G., *Information security risk management: a holistic framework*, 2004, University of Johannesburg.
- [17] Refsdal, A., B. Solhaug, and K. Stølen, *Cyber-risk management*, in *Cyber-Risk Management* 2015, Springer. p. 33-47.
- [18] Zudin, R., *Analysis of information risk management methods*. University of Jyväskylä, 2014.
- [19] Carrillo Sánchez, J.P., *Guía y análisis de gestión de riesgos en la adquisición e implantación de equipamiento y servicios de tecnologías de información y comunicaciones para proyectos de alcance nacional*, 2012, Quito: EPN, 2012.
- [20] Pacheco Pozo, D.C., *Propuesta de un plan de contingencia de TI para la empresa LOGICIEL*, 2016, Quito, 2016.
- [21] Dehkhoda, D., *Combining IRAM2 with Cost-Benefit Analysis for Risk Management: Creating a hybrid method with traditional and economic aspects*, 2018: Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology.
- [22] Wangen, G. *An initial insight into information security risk assessment practices*. in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*. 2016. IEEE.
- [23] Pandey, S. and K. Mustafa, *Risk Assessment Framework (RAF)*. International Journal of Advanced Research in Computer Science, 2010. **1**(3).
- [24] López, D., O. Pastor, and L.J.G. Villalba, *Concepto y Enfoques sobre el Análisis y la Gestión Dinámica del Riesgo en Sistemas de Información*. Actas de la XII Reunión Española de Criptología y Seguridad de la Información (RECSI 2012), Donostia-San Sebastián, España, 2012.
- [25] Saripalli, P. and B. Walters. *Quirc: A quantitative impact and risk assessment framework for cloud security*. in *2010 IEEE 3rd international conference on cloud computing*. 2010. Ieee.
- [26] Li, S., et al., *An improved information security risk assessments method for cyber-physical-social computing and networking*. IEEE Access, 2018. **6**: p. 10311-10319.
- [27] Sicari, S., et al., *A risk assessment methodology for the Internet of Things*. Computer Communications, 2018. **129**: p. 67-79.
- [28] CRAMM. *Siemens Enterprise Communications Ltd. "CRAMM toolkit"*. 2009; Available from: <http://www.cramm.com/>.
- [29] CESG, S.S.a., *CRAMM User Guide*, 2005: Security Service and CESG (UK Government).
- [30] MEHARI, *MEHARI. Risk analysis and treatment Guide*., C.d.I.S.d.I.I. Français, Editor 2010.
- [31] MEHARI, *MEHARI. Overview*., C.d.I.S.d.I.I. Français, Editor 2019.
- [32] OCTAVE. *CERT – Software Engineering Institute, Carnegie Mellon*. 2009; Available from: <http://www.cert.org/octave/>.
- [33] NIST SP 800-30 Rev.1, *NIST SP 800-30 Rev. 1. Guide for Conducting Risk Assessments*., N.I.o.S.a.T. (NIST), Editor 2012, Sep 2012.
- [34] Bjørn, A. *CORAS, A Platform for Risk Analysis on Security Critical Systems—Model-based Risk Analysis Targeting Security*. 2002. EWICS Symposium.
- [35] de la Défense Nationale, S.G., *EBIOS-Expression des Besoins et Identification des Objectifs de Sécurité, Méthode de Gestion des risques*, 2010.
- [36] Telindus, *ISAMM - Information Security Assessment & Monitoring Method*., 2002.
- [37] BSI_S100-2, *BSI-Standard 100-2. IT-Grundsutz Methodology*., F.O.f.I.S. Germany., Editor 2008.
- [38] AS/NZS 4360:2004, *Standards Australia and Standards New Zealand, in Risk Management* 2004, Standards Association Melbourne Info Center: Sydney, NSW.
- [39] ISO_31000:2018, *ISO 31000:2018. Risk Management - Guidelines*, 2018, ISO/TC 262 Risk management: International Organization for Standardization. p. 16.
- [40] Stewart, T.A., *Trying to grasp the intangible*. Fortune, October, 1995. **2**(1995): p. 91-93.

Luis Enrique Sánchez es Doctor y Máster en Informática y es profesor asociado de la Universidad de Castilla-La Mancha (Ciudad Real, España), Master en Auditoría de Sistemas de Información por la Universidad Politécnica de Madrid, y Auditor Certificado de Sistemas de Información por ISACA. Es Director de los departamentos de Servicios Profesionales y de I+D de la empresa Sicaman Nuevas Tecnologías S.L. Sus actividades de investigación son sistemas de seguridad de gestión, métricas de seguridad, minería de datos, limpieza de datos e inteligencia de negocios. Participa en el grupo de investigación GSyA del departamento de sistemas de información y tecnologías de la Universidad de Castilla-La Mancha, en Ciudad Real (España). Su correo electrónico es luisenrique@sanchezcrespo.org.

Antonio Santos-Olmo es Licenciado en Informática y es profesor asociado de la Escuela Superior de Informática de la Universidad de Castilla-La Mancha en Ciudad Real (España), Master en Auditoría de Sistemas de Información por la Universidad Politécnica de Madrid, y Auditor Certificado de Sistemas de Información por ISACA. Es Director de los departamentos de Software Factory de la empresa Sicaman Nuevas Tecnologías S.L. Sus actividades de investigación son sistemas de seguridad de gestión, métricas de seguridad, minería de datos, limpieza de datos e inteligencia de negocio. Participa en el grupo de investigación GSyA del Departamento de de sistemas de información y tecnologías de la Universidad de Castilla-La Mancha, en Ciudad Real (España). Su correo electrónico es asolmo@sicaman-nt.com.

Julio Moreno es Máster y estudiante de doctorado en Informática por la Universidad de Castilla-La Mancha. Su investigación se centra en la seguridad y privacidad de los datos, así como en la creación de un marco de gobierno para Big Data. Es miembro del grupo de investigación GSyA del departamento de sistemas de información y tecnologías de la Universidad de Castilla-La Mancha. Su correo electrónico es julio.moreno@uclm.es.

Manuel A. Serrano es Máster y doctor en Informática por la Universidad de Castilla-La Mancha. Es profesor titular en la Escuela Superior de Informática de la Universidad de Castilla-La Mancha en Ciudad Real. Su investigación se centra en la calidad de software de los datos, la medición de software y la calidad y medición de almacenes de datos y Big Data. Su correo electrónico es manuel.serrano@uclm.es.

David G. Rosado tiene un Máster y es doctor en Informática por la Universidad de Málaga y por la Universidad de Castilla-La Mancha, respectivamente. Es profesor titular en la Escuela Superior de Informática de la Universidad de Castilla-La Mancha en Ciudad Real. Su actividad investigadora se centra en seguridad de sistemas de información, Cloud Computing y Big Data. Sobre estos temas, él ha publicado numerosos artículos en conderencias de ámbito nacional e internacional, también es editor y coeditor de varios libros. Es autor de varias publicaciones en revistas nacionales e internacionales (Information Software Technology, System Architecture, Network and Computer Applications, etc.). Él es miembro del comité de programa de numerosas conferencias y workshops nacionales e internacionales. Es miembro del grupo de investigación GSyA del departamento de sistemas de información y tecnologías de la Universidad de Castilla-La Mancha. Su correo electrónico es david.grosado@uclm.es.

Eduardo Fernández-Medina es Máster y doctor en Informática por la Universidad de Castilla-La Mancha. Es profesor catedrático en la Escuela Superior de Informática de la Universidad de Castilla-La Mancha en Ciudad Real (España). Su actividad investigadora se centra en el campo de la seguridad de sistemas de información, en particular en seguridad en Big Data, Cloud Computing y sistemas ciberfísicos. En estas temáticas, él es coeditor de varios libros y capítulos de libros, y ha publicado numerosos artículos en conferencias nacionales e internacionales (BPM, UML, ER, ESORICS, TRUSTBUS, etc.). Es autor de más de cincuenta publicaciones en revistas internacionales (Decision Support Systems, Information Systems, ACM Sigmod Record, Information Software Technology, Computer & Security, Computer Standards and Interfaces, etc.). Él lidera el grupo de investigación GSyA del departamento de sistemas de información y tecnologías de la Universidad de Castilla-La Mancha y pertenece a varias asociaciones profesionales y de investigación (ATI, AEC, AENOR, etc.). Su correo electrónico es eduardo.fdezmedina@uclm.es.

Inteligencia Artificial aplicada al Análisis Forense Digital: Una revisión preliminar

J. J. Cano, J. D. Miranda y S. Pinzón

Resumen—The digital forensic analysis seeks the application of scientific and statistical techniques to identify, collect, preserve and present the relevant digital evidence that allows the hypothesis to be affirmed or rejected against a possible criminal act. The current methods of digital forensic analysis, although effective for the visual analysis of the material evidence, do not allow to execute in an automated way and for large volumes of data, correlation studies on the obtained files, validation of metadata and identification of anomalies in files of text, graphics or audiovisual. It is for this reason that artificial intelligence techniques have been proposed for data processing, identifying patterns and trends that allow noticing aspects that are not visually perceptible. This paper discusses the role that artificial intelligence can play in digital forensic analysis, proposing a review of the literature, in order to illustrate the areas of computer forensics in which artificial intelligence techniques have been used to date. This, to identify a new work niche in this area, hoping that the ideas in this document can represent promising directions for the development of more efficient and effective computer forensic tools.

Palabras clave — *Forensic Computing, Forensic Analysis, Artificial Intelligence, Vector Support Machines, Artificial Neural Networks, Autonomous Systems, Intelligent Agents.*

I. INTRODUCCIÓN

Debido al rápido crecimiento en la aplicación de tecnologías digitales en diferentes entornos de la sociedad, son cada vez más los casos de ciberdelincuencia y mayor la cantidad de información que se logra extraer como rastro de un acto de delincuencia digital.

En complemento, los ataques informáticos se han vuelto más sofisticados, no solo por la experticia de los atacantes, sino por las herramientas que usan estos para acometer las acciones no autorizadas. En consecuencia, se hace necesario el uso de instrumentos que asistan al personal especializado para el estudio de los elementos materiales probatorios recopilados frente a la ocurrencia de un evento de carácter informático y

que favorezcan en términos de efectividad y velocidad a la resolución de los diferentes casos forenses.

Sin embargo, a través de los años la intervención humana ha sido en algunos casos insuficiente para el análisis oportuno de un ataque y su consiguiente respuesta. Estos ataques son encabezados por atacantes habilidosos que emplean herramientas cada vez más sigilosas y perfeccionadas para esta labor [1], tales como agentes inteligentes, gusanos o virus informáticos, que son analizados por personal humano que desconoce su comportamiento y que busca de forma superficial rastros de una dinámica no conocida subyacente. Adicionalmente, existen otros desafíos, tal como lo documenta [2], que hacen que el análisis forense digital sea complejo, tedioso y en ocasiones infructuoso:

- La complejidad del problema y la heterogeneidad de los datos en su adquisición.
- El volumen excesivo de datos procedentes de múltiples fuentes y la falta de técnicas estandarizadas para procesarlos.
- La falta de técnicas que encuentren la correlación en la información contenida en la evidencia digital.
- La falta de estandarización en las zonas horarias y los registros de tiempo de los eventos encontrados.

Estos factores hacen que se requiera la aplicación de extensas etapas adicionales de preprocesamiento de los datos, se aumente el tiempo de respuesta y no se termine de efectuar un proceso eficiente de análisis forense, dejando al atacante en ocasiones dentro del sistema, lo que aumenta los riesgos para la empresa, los costos en reparaciones y la recuperación de los daños informáticos.

Un reporte del Ponemon Institute [3] revela que, a 2014, se ha aumentado el costo del cibercrimen en más del 9% con respecto a años anteriores, y el tiempo para resolver un ataque a incrementado a 45 días, lo que representa un alza de 40% respecto a las mediciones de los años previos. Esto significa que las técnicas actuales utilizadas para el análisis forense digital no son suficientes para el análisis oportuno y concluyente de los eventos informáticos que se presentan.

Es por esta razón que es propicio el uso adicional de agentes semiautónomos inteligentes que puedan aportar en la eficacia del análisis forense y en la toma de decisiones, con base en la experiencia. Esto puede hacerse mediante la implementación de métodos de Inteligencia Artificial (AI) como las Máquinas de

J. J. Cano, Facultad de Ingeniería de Sistemas e Informática, Universidad Pontificia Bolivariana de Bucaramanga. Colombia. jjcano@yahoo.com.

J. D. Miranda, Facultad de Electrónica e Ingeniería de Sistemas e Informática, Universidad Pontificia Bolivariana de Bucaramanga. Colombia. juliandariomiranda@gmail.com. Corresponding author.

S. Pinzón. Facultad de Ingeniería de Sistemas e Informática, Universidad Pontificia Bolivariana de Bucaramanga. Colombia. spinzonsarmiento@gmail.com.

Soportes Vectoriales (SVM), Redes Neuronales (NN), Agentes Inteligentes y Aprendizaje de Máquina (ML), entre otros, aplicados a los sistemas que asisten al personal especializado en la detección, prevención y mitigación del cibercrimen [4].

En el campo del análisis forense automatizado se utilizan métodos de AI con el objetivo de automatizar los diferentes procesos y análisis que se realizan en los dispositivos, teniendo en cuenta un amplio volumen de datos. Además, estos datos son inocuos por sí mismos, ya que provienen de diferentes fuentes y para ser útiles debe eliminarse la alta correlación existente, con el fin de descartar los que son intrascendentes o similares entre sí, y ejecutar un análisis pertinente de la dinámica no visible de estos.

Tal como lo documenta Laurance Merkle [5], las herramientas existentes en el mercado que realizan análisis estadísticos, análisis de tráfico de red y análisis de sesiones, entre otras, lo hacen de forma superficial y mediante el estudio de umbrales, razón por la cual resultan insuficientes para procedimientos de análisis forense multinomial (múltiples variables de entrada).

El objetivo de este artículo es presentar los resultados de un breve estudio del estado del arte de referentes de interés en el campo de la implementación de técnicas de inteligencia artificial, aprendizaje automático, máquinas de soportes vectoriales y sistemas expertos, puestos a disposición del análisis forense digital, de tal forma que, se refuerce la identificación de patrones y componentes que no son visualmente perceptibles, se correlacionen eventos para encontrar una secuencia no evidente y se puedan tomar decisiones oportunas frente a los eventos informáticos. A continuación, se detalla la fundamentación teórica que enmarca el contexto de esta investigación, seguida de la metodología propuesta para la búsqueda de información, los resultados obtenidos y la discusión de estos.

II. FUNDAMENTACIÓN TEÓRICA

Esta sección se divide en cinco temáticas de estudio, consideradas relevantes para el desarrollo de la investigación: inteligencia artificial, aprendizaje automático, máquinas de soportes vectoriales, redes neuronales artificiales, y sistemas expertos y agentes inteligentes.

A. Inteligencia Artificial

La Inteligencia Artificial (AI) está definida como la capacidad de un sistema para interpretar correctamente datos externos, para aprender de dichos datos y emplear esos conocimientos para lograr tareas y metas concretas a través de la adaptación dinámica y flexible [6]. Según [7], una máquina es considerada como inteligente cuando se trata de un agente dinámico que percibe su entorno y lleva a cabo acciones que maximicen sus posibilidades de éxito en algún objetivo o tarea. En otras palabras, la AI involucra un agente dinámico flexible que aprenda de las observaciones con el fin de presentar una respuesta cada vez más acertada y acorde con el aprendizaje.

Son muchos los campos en los que se aplica AI, entre los cuales destacan la lingüística computacional, la medicina, la robótica, los videojuegos, la domótica, la automatización y la seguridad informática; buscando resolver problemas relacionados con la búsqueda de nuevas heurísticas, la representación del conocimiento, la planeación de estrategias, el procesamiento del lenguaje, la percepción de los patrones y el análisis forense. En este último contexto, la AI ha puesto especial atención en el aprendizaje de máquina, tal como lo muestra la Fig. 1, una ramificación de la AI en la que se busca el desarrollo de técnicas estadísticas que le permiten a la máquina mejorar la ejecución de las tareas con la experiencia.

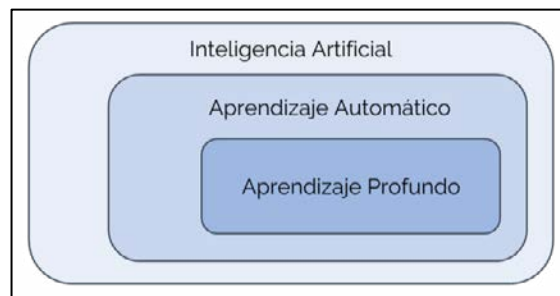


Fig. 1. Relación entre la inteligencia artificial, el aprendizaje de máquina y el aprendizaje profundo. Elaboración propia.

Dentro del aprendizaje de máquina existen desarrollos algorítmicos específicos que permiten que el software se entrene a sí mismo para llevar a cabo tareas como el habla y el reconocimiento de patrones, mediante la exposición a gran cantidad de datos, en lo que se conoce como el aprendizaje profundo.

De esta forma, para resolver problemas mediante métodos computacionales, se requiere de secuencias de instrucciones que transformen un conjunto de variables de entrada conocidas, en un grupo de variables de salida deseadas. Para procesos en los cuales no se conoce su comportamiento en su totalidad, se construye una aproximación. Ésta puede no explicar la totalidad de los eventos consultados, pero permite la detección de patrones que sirven como indicios para su identificación y clasificación. Este es el nicho del Aprendizaje Automático, los sistemas expertos y agentes inteligentes.

B. Aprendizaje Automático

En general, el aprendizaje automático es el proceso de programación de una unidad computacional para optimizar los criterios de rendimiento, haciendo uso de muestras recolectadas de experiencias pasadas [8]. Un modelo de aprendizaje puede ser supervisado o no supervisado. En el aprendizaje supervisado se orientan las predicciones del algoritmo mediante un conjunto de categorías o etiquetas. Los algoritmos de clasificación y regresión son ejemplos de aprendizaje supervisado. En la clasificación se intenta hacer una predicción de la categoría a la que corresponde una observación mientras que, en regresión, se intenta hacer una predicción de un valor numérico.

Por otro lado, en el aprendizaje no supervisado no existen categorías adjuntas a los datos de entrada. Los algoritmos de Clustering, estimación de densidad y reducción de dimensionalidad son ejemplos de aprendizaje no supervisado. En Clustering se intenta agrupar datos con características similares, mientras que en estimación de densidad se pretende encontrar valores estadísticos que describan al conjunto de datos. En cuanto a la reducción de dimensionalidad, se busca reducir la cantidad de características de los datos para reducir el costo computacional de su procesamiento y posibilitar su representación más fácilmente [9].

Para que el modelo aprenda de los datos se ejecuta un proceso inicial de entrenamiento que se realiza de forma iterativa. Cuando el modelo de aprendizaje ha sido expuesto a todo el conjunto de entrenamiento, se dice que ha pasado un *epoch* [10]. Existen tres conjuntos de datos con los que se ejecutan las fases de entrenamiento, validación y prueba: el training set es el conjunto de datos usados como referencias para seleccionar los pesos asociados a las unidades de procesamiento del modelo de aprendizaje y crear las conexiones entre las mismas; el *validation set* corresponde al conjunto de datos usados durante el proceso de entrenamiento para contrastar el desempeño por cada ciclo de aprendizaje, clasificando datos desconocidos; y el test set es el conjunto de datos desconocidos para el modelo de aprendizaje, con los que se prueba el desempeño.

C. Máquinas de Soportes Vectoriales (SVM)

Las máquinas de soportes vectoriales (SVM) son una familia de algoritmos de aprendizaje automático de tipo supervisado que generalmente se emplea en procesos de clasificación y regresión de dos conjuntos de datos. Estos algoritmos han sido aplicados ampliamente en áreas como el reconocimiento de escritura, la detección de rostros, categorización de texto, entre otras. El método combina el aprendizaje estadístico y la optimización convexa, combinando una máquina de soporte vectorial para clasificación y una para regresión. Contiene una base de conocimiento o ejemplos de entrenamiento que generalmente pertenecen a una de dos categorías, siendo esta una clasificación de tipo binaria [11].

El nombre se deriva del conjunto de puntos de datos o *vectores de soporte* que contienen la información. Una SVM es básicamente una máquina de aprendizaje lineal diseñada para resolver problemas de clasificación usando el principio de separación de clases. El objetivo es encontrar un hiperplano, un plano en un espacio multidimensional de separación lineal que separe dos clases de interés. El hiperplano se ubica entre clases para cumplir dos condiciones: que todos los vectores de datos que pertenezcan a una misma clase se ubiquen del mismo lado del hiperplano y maximizar la distancia entre los vectores de datos más cercanos en ambas clases [11].

La función del SVM es la de encontrar la mejor línea, plano o hiperplano que divida el grupo de datos en dos clases, utilizando un kernel (núcleo) que se compone de un conjunto

de funciones matemáticas para transformar los datos de entrada en la forma deseada. En ese sentido, un hiperplano óptimo es aquel que deja el margen máximo entre las dos clases de salida, siendo el margen la distancia entre el hiperplano y el dato más cercano de cualquiera de los conjuntos.

Las máquinas SVM tiene un buen comportamiento para conjuntos de datos pequeños, pero al aumentar la cantidad de datos es más difícil encontrar una posición del hiperplano que divida adecuadamente las dos clases de datos, además del incremento de carga computacional y aumento de la complejidad en el aprendizaje conforme aumenta el número de muestras [12].

D. Redes Neuronales Artificiales (ANN)

Las redes neuronales artificiales (ANN: Artificial Neural Networks) se plantean como una alternativa computacional para la toma de decisiones dentro del dominio del Aprendizaje Automático. Esta técnica busca emular la capacidad de aprendizaje natural de los seres vivos, la cual es atribuida al sistema neuronal de su cerebro.

La unidad básica de procesamiento de una red neuronal artificial es la neurona. Las redes neuronales funcionan mediante la interacción de conjuntos de neuronas con características diferentes. Dichas características son explicadas a continuación:

- Función de propagación: cada neurona tiene una serie de entradas provenientes de otras neuronas. Estas señales de entrada son atenuadas o amplificadas por un factor de peso y son operadas en conjunto por una función de propagación que es comúnmente una suma ponderada.
- Función de activación: es la función de umbral que determina la acción de una neurona, dependiendo del valor de entrada proveniente de la función de propagación [8].
- Función de salida: es la función encargada de calcular el valor de salida de la neurona para ser transferido como entrada a otras neuronas [13].

La topología de una red neuronal está ligada con el algoritmo de aprendizaje usado para entrenar la red. Los factores que definen la topología de la red son las capas y la naturaleza de las conexiones entre las neuronas. En una red neuronal, una capa es un conjunto de nodos (pueden ser neuronas o fuentes de datos) con características similares, subdivididas en capas de entrada, capas ocultas y capas de salida. Las conexiones entre las neuronas y las capas pueden ser unidireccionales (feedforward) o recurrentes (con al menos un lazo de realimentación o feedback).

La capacidad de aprendizaje mediante un proceso de entrenamiento y la aplicación de la experiencia adquirida en este proceso le otorga a las ANN la capacidad de responder apropiadamente a situaciones a las que no había sido expuesta

[14]. Los sistemas de aprendizaje cambian sus características para poder adaptarse al problema que se está afrontando y conseguir la generalización de la comprensión del problema. Las ANN pueden aprender mediante el desarrollo de nuevas conexiones, del cambio de la ponderación de sus conexiones, la creación de nuevas neuronas y el cambio los valores de umbral en la función de activación, entre otros.

E. Sistemas expertos y agentes inteligentes

Los sistemas inteligentes consisten en softwares de respuesta a preguntas en un determinado dominio, emulando la capacidad de toma de decisiones de un experto humano. Son utilizados generalmente para apoyar la decisión de un experto en diferentes áreas del conocimiento. El sistema contiene una base de conocimientos sobre un dominio específico y se conecta con un motor de inferencia para la derivación de las respuestas basadas en este conocimiento [15].

Los agentes inteligentes son componentes de software que poseen características de comprensión de un lenguaje, capacidad para tomar decisiones y actuar según corresponda, usualmente usados para la defensa contra ataques frecuentes. Junto con un sistema de redes neuronales, estos agentes pueden conformar un método de detección eficaz [16].

III. METODOLOGÍA

El análisis del estado del arte concerniente con la aplicación de inteligencia artificial en el análisis forense digital se hizo mediante consultas a la base de datos de Scopus de Elsevier, utilizando la siguiente sentencia de búsqueda:

TITLE-ABS-KEY ("Digital Forensics" AND "Forensic Analysis" AND ("intelligent agent" OR "Support Vector Machines" OR "Artificial Neural Networks")) AND (LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA, "ENGI"))

En esta consulta se hizo una búsqueda todos aquellos documentos que contuvieran las frases "Forense Digital" y "Análisis Forense", y cualquiera de las técnicas de inteligencia artificial: "Redes Neuronales Artificiales", "Máquinas de Soportes Vectoriales" o "Agentes Inteligentes". La búsqueda se enmarca en las ciencias de la computación y la ingeniería, con el fin de obtener aplicaciones de la inteligencia artificial dentro de la computación, tal que muestre resultados preliminares algorítmicos y de desempeño de modelos de aprendizaje. La consulta se hizo para documentos publicados a partir del año 2000 en adelante, excluyendo patentes o cualquier documento no publicado.

La temática analizada en esta investigación es de carácter novedoso y se encuentra poca literatura disponible y un número limitado de artículos que detallen los modelos, propuestas y resultados preliminares de desarrollo e implementación. Inicialmente fueron encontrados 161 documentos que involucraban la temática de forense digital, de los cuales 17 hacían énfasis en el análisis forense y el uso de técnicas de

inteligencia artificial, siendo seleccionados 11 que documentaban puntualmente la información requerida: técnicas implementadas preliminarmente, tamaño de las observaciones de entrenamiento, validación y prueba de la técnica, y desempeño del modelo implementado. En la siguiente sección se detallan los resultados del análisis de referentes ejecutado.

IV. RESULTADOS DEL ANÁLISIS

Un breve estudio del estado del arte fue ejecutado con el fin de identificar las diversas aplicaciones de la Inteligencia Artificial al análisis forense digital, haciendo especial énfasis en el uso de técnicas de Aprendizaje Automático, entre las que se encuentran las Máquinas de Soportes Vectoriales y las Redes Neuronales Artificiales, y el uso de sistemas expertos y agentes inteligentes.

En la Tabla I se ha detallado el estado del arte relevante para la investigación. De esta tabla se puede observar que se han desarrollado diversos trabajos en el ámbito del análisis forense digital que incluyen: la detección de alteraciones a imágenes por medios digitales, la identificación de atributos y características físicas y químicas de objetos y la determinación del contenido de una imagen forense para su análisis automatizado, todos enmarcados en el uso de técnicas de aprendizaje automático e inteligencia artificial. A continuación, se detallan los trabajos relevantes para la investigación por temática.

A. Máquinas de Soportes Vectoriales

Algunos expertos han aplicado técnicas de aprendizaje automático en informática forense, en particular máquinas de soportes vectoriales (SVM), que se aplican generalmente al reconocimiento de escritura, detección de rostros, procesamiento del lenguaje natural y visión artificial. Las SVM pertenecen a una clase de algoritmos de Machine Learning denominados métodos kernel que se usan continuamente en la detección de intrusiones, debido a su alta velocidad de aprendizaje y escalabilidad. A continuación, se detalla el proceso y resultados de autores que han implementado estas técnicas para ejecutar un análisis forense, enmarcado en distintas áreas de estudio y con diversos objetivos.

S. Mukkamala y A. Sung [17] estudiaron en 2003 la implementación de SVM para el análisis forense de redes de datos, sobre una base de datos de 494,021 eventos procedentes del ataque de DARPA en 1999, considerando cuatro tipos de ataques: análisis de vulnerabilidades invasivo, denegación de servicios (DOS), acceso no autorizado con privilegios de súper usuario (U2Su) y acceso no autorizado desde máquina remota (R2L). El desempeño del clasificador por SVM logra una precisión ligeramente superior al 99% para las cuatro clases de ataques evaluadas.

Por otro lado, A. Mikkilineni, et. al. [18] en 2005, implementaron una clasificación multiclase utilizando SVM, para la identificación de impresoras a partir de documentos impresos, considerando el tamaño de letra, el tipo de fuente, el tipo de papel y la edad de la impresión. Se consideraron 5,000 datos de entrenamiento y prueba con 10 clases de impresoras de

TABLA I. ESTADO DEL ARTE DE INTERÉS DEL USO DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL EN ANÁLISIS FORENSE.

Método	Ref	Objetivo	Dataset de observaciones	Desempeño
Máquinas de Soportes Vectoriales	[17]	Análisis forense de redes de datos considerando cuatro tipos de ataques: análisis de vulnerabilidades invasivo, denegación de servicios (DOS), acceso no autorizado con privilegios de súper usuario (U2Su) y acceso no autorizado desde máquina remota (R2L).	Entrenamiento (50%): 247,010 Prueba (50%): 247,010 Total dataset: 494,021	99.5% de precisión
	[18]	Identificación de impresoras a partir de documentos impresos, considerando el tamaño de letra, el tipo de fuente, el tipo de papel y la edad de la impresión.	Entrenamiento (50%): 5,000 Prueba (50%): 5,000 Total dataset: 10,000	84% al 93% de precisión
	[19]	Determinación del tipo de archivos contenidos en el material probatorio de una investigación.	Entrenamiento (90%): 3,240 Prueba (10%): 360 Total dataset: 3,600	81% al 98% de precisión
	[20]	Desarrollo de un modelo de aprendizaje para la identificación de modificaciones de filtrado y re-muestreo en imágenes digitales.	Entrenamiento (25%): 1,250 Prueba (75%): 3,250 Total dataset: 5,000	99.35% de precisión
	[21]	Desarrollo de un modelo para la clasificación de bloques de archivos con formatos conocidos.	Entrenamiento (20%): 50,400 Prueba (80%): 201,600 Total dataset: 252,000	85% al 95% de precisión
Redes Neuronales Artificiales	[23]	Identificación de clases de vidrio con base en atributos químicos (índice de refracción, cantidad de sodio, magnesio, aluminio, silicio, potasio, calcio, bario, hierro y tipo de vidrio) de muestras encontradas en las escenas donde ocurrieron los eventos criminales.	Entrenamiento (80%): 192 Prueba (20%): 38 Total dataset: 240	73% al 85% de precisión
	[24]	Identificación de atributos espaciales en imágenes con esteganografía LSB con cinco densidades de contenido embebido (payload): 0.1, 0.2, 0.3, 0.4 y 0.5 bpp (bits por pixel).	Entrenamiento (70%): 7,000 Prueba (30%): 3,000 Total dataset: 10,000	84% al 86% de precisión
	[25]	Desarrollo de un modelo para la detección de posible manipulación de imágenes digitales a color.	Entrenamiento (80%): 160,000 Prueba (20%): 40,000 Total dataset: 200,000	95% de precisión
	[26]	Análisis forense de imágenes JPEG para notar transformaciones espaciales.	Entrenamiento (65%): 284,000 Prueba (35%): 150,000 Total dataset: 434,000	84% al 99% de precisión
	[27]	Desarrollo de un algoritmo para la detección y localización de falsificaciones de imágenes mediante funciones de remuestreo y aprendizaje profundo.	Entrenamiento (80%): 100,000 Prueba (20%): 25,000 Total dataset: 125,000	95% de precisión
Sistemas expertos y agentes inteligentes	[28]	Desarrollo de una propuesta de software MADIK para la investigación digital multi agente que permita suplir la dificultad que tiene un experto de determinar de forma rápida qué evidencia es relevante cuando se analiza un crimen.		De 69% a 74% de cubrimiento

salida para la clasificación, obteniendo un desempeño en la clasificación de entre el 84% y el 93% de precisión.

Con el fin de determinar el tipo de archivos contenidos en el material probatorio de una investigación, Q. Li y A. Ong [19] en 2010 clasificaron los fragmentos de archivos utilizando SVM, analizando cinco clases de salida: archivos JPEG (800 ejemplos), archivos MP3 (800 ejemplos), archivos PDF (800 ejemplos), archivos DLL (800 ejemplos) y archivos binarios ejecutables de Windows (400 ejemplos). La distribución entre entrenamiento y prueba fue del 90% y 10%, respectivamente, obteniendo un desempeño del clasificador medido en precisión en el intervalo de 81% a 98%.

Identificar modificaciones ocultas en imágenes o develar alteraciones en el formato de los archivos son otros campos trabajados en el análisis forense digital, pues puede existir contenido embebido no visible en las imágenes que pueda ser de utilidad, o los archivos pueden estar particionados para evitar mostrar su contenido al abrirlas, haciendo que parezcan dañados.

En este primer campo, Cai, et. al. [20] desarrollaron un modelo de SVM que permite identificar si una imagen ha sido alterada por medios digitales, incluyendo operaciones como submuestreo, compresión, filtrado de mediana y filtrado de media. Esto, a partir de una base de datos de 1,000 imágenes a

color por cada operación de modificación, divididas en 25% para entrenamiento y 75% para pruebas del modelo. El desempeño del clasificador multiclase fue medido en precisión y alcanzó el 99.35%.

De forma complementaria, Sportiello y Zanero [21] desarrollaron un modelo de SVM que, mediante descriptores frecuenciales, permite clasificar los bloques de archivos que han sido separados o malformados, haciendo énfasis en los formatos: bmp, doc, exe, gif, jpg, mp3, odt, pdf y ppt. Esto, a partir de una base de datos heterogénea de archivos en estos formatos que fueron seccionados en bloques de 512 bytes para un total de 252,000 bloques que fueron analizados para el entrenamiento (20% de las observaciones) y la prueba (80% de las observaciones) del modelo. El desempeño final del mismo medido en precisión alcanza un máximo de 95% con un mínimo de clasificación de 85%.

B. Redes Neuronales Artificiales

Una segunda técnica utilizada son las redes neuronales artificiales, consisten en una colección de elementos que están interconectados y se transforman en un conjunto de salidas deseadas. La red neuronal realiza un análisis de la información y proporciona una estimación de probabilidad que coincide con los datos para los cuales ha sido entrenada para reconocer. La

red neuronal va obteniendo eficacia al irse entrenando el sistema con la entrada y la salida del problema deseado. A partir de ese entrenamiento se obtienen mejores resultados y la configuración de la red se refina conforme pasa el tiempo [17].

R. Mohammad [22] ha demostrado que es posible identificar si el sistema operativo ha sufrido modificaciones no autorizadas a partir de la implementación de un modelo de red neuronal.

A. Tallón-Ballesteros y J. Riquelme [23] en 2014, por otro lado, han aplicado métodos de clasificación multiclase usando redes bayesianas, árboles de decisión y redes neuronales artificiales, para la identificación de clases de vidrio con base en atributos químicos (índice de refracción, cantidad de sodio, magnesio, aluminio, silicio, potasio, calcio, bario, hierro y tipo de vidrio) de muestras encontradas en las escenas donde ocurrieron los eventos criminales, a partir de una base de datos del USA Forensic Science Service con 240 observaciones. Los algoritmos implementados fueron validados haciendo uso de validaciones cruzadas de 4 folds, es decir, subdividiendo el conjunto de datos de prueba en cuatro subconjuntos homogéneos y ejecutando las pruebas pertinentes sobre los datos. Los autores obtienen un desempeño en el mejor clasificador (el de redes neuronales) de entre 73% y 75%.

Qian, et. al. [24] en 2016, desarrollaron un modelo de aprendizaje basado en Redes Neuronales Convolucionales (CNN) para la identificación de atributos espaciales en imágenes con esteganografía LSB, a partir de imágenes procedentes de la base de datos BOSSbase (Break Our Steganography System) v1.01, una base de datos de imágenes a escala de grises diseñada para ejecutar pruebas de estegoanálisis. Esta base de datos contiene 10,000 imágenes portadoras y con esteganografía con una distribución del 50% y con resolución espacial de 512x512. El 70% de este conjunto de datos fue asignado para entrenamiento del modelo, el 10% para validación y el 20% para pruebas. Se tuvieron en cuenta cinco densidades de contenido embebido (payload): 0.1, 0.2, 0.3, 0.4 y 0.5 bpp (bits por píxel). El clasificador logró un desempeño medido en precisión de entre el 84% y 86%.

D. Kim y H. Lee [25] en 2017 han utilizado un enfoque de redes neuronales convolucionales, un tipo especializado de redes neuronales que permite aprender de una mayor cantidad de información, almacenando conocimiento de los entrenamientos. Esta característica les permite a los autores desarrollar un modelo para la detección de posible manipulación de imágenes digitales a color. El modelo fue entrenado con ejemplos gráficos procedentes de una base de datos estandarizada con dimensión espacial 256x256, que contienen cuatro tipos de ruido, procesando las imágenes con filtros de mediana y Gaussianos. El modelo fue entrenado con 160,000 imágenes (80% del dataset) y probado con 40,000 imágenes (20% del dataset), consiguiendo un desempeño del detector de 95% de precisión.

De forma similar, N. Bonettini [26] en 2017 ha usado un enfoque de redes neuronales convolucionales para el análisis forense de imágenes JPEG, teniendo en cuenta una base de datos RAISE de imágenes RAW de 284,000 ejemplos que fue transformada en imágenes a escala de grises que fueron comprimidas, con resoluciones espaciales variadas equitativamente desde 64x64 hasta 256x256. Para el

entrenamiento, el 70% del dataset original fue tenido en cuenta, con un tamaño de batch de 128 (cantidad de ejemplos por subconjunto de entrenamiento) y 30 epochs de entrenamiento (cantidad de iteraciones por unidad de entrenamiento). La validación del algoritmo fue ejecutada con el 30% del dataset y la prueba con 150,000 imágenes independientes del conjunto de entrenamiento y validación. El desempeño del clasificador está entre el 84% y 99% de precisión.

J. Bunk et. al. [27] en 2017, desarrollaron un algoritmo para la detección y localización de falsificaciones de imágenes mediante funciones de remuestreo y aprendizaje profundo, usando redes neuronales LSTM, una tipología de redes neuronales especializada en almacenar memoria de corto y largo plazo en cada iteración de entrenamiento, haciendo que sea más rápido el aprendizaje con una base de conocimiento existente. Para esto, tuvieron en cuenta un dataset de 100,000 ejemplos de las bases de datos de UCID y RAISE, y para la prueba la base de datos de NIST Nimble 2016 con tres clases de modificaciones en las imágenes: copiado y clonado, remoción y sobreposición. Los autores obtienen un 95% de precisión con el modelo de LSTM ANN en la detección de falsificaciones.

C. Sistemas expertos y agentes inteligentes

Una propuesta de software presentada en el escenario internacional es la de MADIK, un Toolkit para la investigación digital multi agente (de sus siglas en inglés Multi-Agent Digital Investigation ToolKit) elaborado por B. Hoelz, et. al. [28] en 2008. Esta es una apuesta que se ha sido desarrollada con el fin de suplir la dificultad que tiene un experto de determinar de forma rápida qué evidencia es relevante cuando se analiza un crimen. Ante la falta de herramientas que colaboren en el preanálisis de las evidencias y su correlación, nace este marco de trabajo basado en un conjunto de herramientas usando un sistema de multi agente. Es implementado usando JADE, marco de trabajo basado en lenguaje Java muy común en el desarrollo de sistemas multi agente.

La arquitectura definida (ver Fig. 2) se divide en roles distribuidos en cuatro niveles de agentes. Se definen agentes autónomos especializados cada uno con distintos objetivos, que pueden colaborar con el trabajo de otros agentes en un único espacio denominado *Blackboard*. Un sistema multi agente permite un uso de los recursos más eficiente y a los agentes operar de forma autónoma en diferentes máquinas y entornos.

La arquitectura de cuatro capas incluye un nivel operativo donde encontramos a los agentes especializados, y unos niveles superiores en donde se encuentran los agentes de gestión, encargados de las decisiones estratégicas y tácticas. Estos últimos están encargados a su vez de la distribución y coordinación de las tareas que ejecutarán los agentes especializados de la capa operacional. La comunicación es jerárquica, ya que los agentes especializados se comunican solo con el gerente operacional, el gerente operacional con el gerente táctico y el gerente táctico con el gerente estratégico.

La plataforma se puede distribuir entre máquinas diferentes y la configuración se realiza desde una interfaz GUI remota. Dentro de las opciones principales que permite la configuración

están las de definir tiempos donde se ejecutan procesos y mover a los agentes de una máquina a otra.

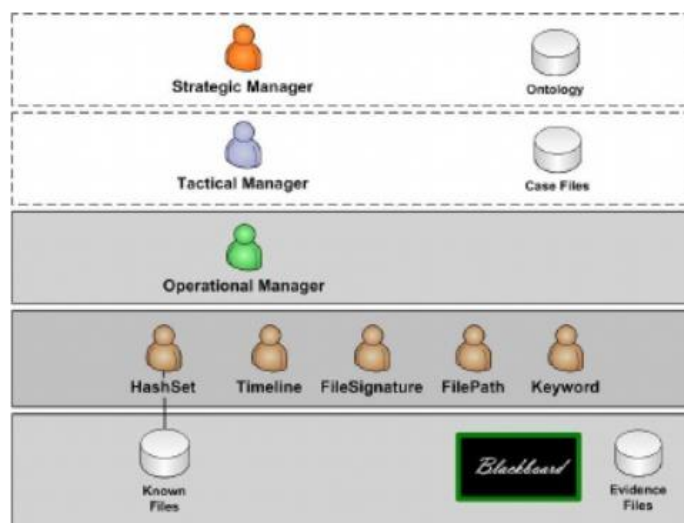


Fig. 2. Arquitectura MADIK. Tomado de: [28].

En el nivel operativo se encuentra la mayor cantidad de agentes desplegados, entre los que se encuentran:

- HashSetAgent: calcula el hash MD5 de un archivo y lo compara con su base de conocimiento.
- FilePathAgent: mantiene en su base de conocimiento carpetas que son usadas habitualmente por aplicaciones que pueden ser de interés para la investigación.
- FileSignatureAgent: analiza el encabezado del archivo para determinar su extensión. Usado para determinar si el atacante cambia la extensión de los archivos.
- TimelineAgent: Examina fecha de creación, de acceso y modificación de los archivos.

Después de varios experimentos realizados con las herramientas en múltiples casos, los autores pueden determinar buenos resultados al compararlos con las herramientas disponibles evaluadas. En comparación con el trabajo realizado por un humano experto, los autores observaron un menor factor de reducción para el sistema de múltiples agentes, aunque en términos del tiempo requerido observaron una reducción de tiempo hasta por seis veces, considerando el mismo contenido examinado por parte humana y parte artificial. El factor de cobertura alcanzó el 80%, una buena cifra, según los autores, que se puede mejorar con bases de conocimientos robustas y con el uso de más agentes especializados.

D. Otras aplicaciones

La Inteligencia Artificial ha sido usada en otros campos en los que se puede extrapolar un estudio forense, como lo son: el reconocimiento de patrones en línea [29] [30], el descubrimiento de conocimiento [31], la computación forense sobre Internet de las Cosas (IoT) [32], la correlación de eventos almacenados en los elementos materiales probatorios [33], entre otras aplicaciones.

El reconocimiento de patrones es un campo particularmente utilizado en el análisis forense digital de elementos materiales probatorios, pues permite identificar grupos de datos y dinámicas no visibles. Es ampliamente utilizado en informática forense para el reconocimiento de patrones en imágenes, tal como se ha precisado, en el que una sección de software realiza la identificación de componentes espaciales y frecuenciales de una imagen [34] [35]. Otra forma de reconocimiento de patrones se aplica en mensajes de correo electrónico para encontrar aquellos con SPAM o que utilicen phishing [36] [37]. También se puede utilizar para reconocimiento de patrones de audio en pistas o sectores de disco [38].

Un tercer campo de la AI que se puede aprovechar en el ámbito forense es la minería de datos y descubrimiento de conocimiento en bases de datos. Aunque ambos conceptos son diferentes, se usan para referirse al mismo procedimiento de recolección de grandes cantidades de datos [31].

V. DISCUSIÓN

Las técnicas actuales de análisis forense digital están basadas en la identificación de variaciones de variables individuales que permiten una identificación de posibles actividades irregulares mediante métodos manuales y visuales. Sin embargo, el exceso en el volumen de datos que deben ser analizados, los patrones que no son identificables al aplicar técnicas manuales y visuales, y las diversas dinámicas de texto, gráficas y audiovisuales que deben ser analizadas, hacen que la complejidad temporal y espacial de los algoritmos crezca, aumentando el tiempo requerido para ejecutar el análisis forense y limitando las capacidades computacionales en la búsqueda de información concluyente. Para esto se han desarrollado técnicas de inteligencia artificial que permiten entrenar un modelo que identifique de forma correlacionada e integrada la actividad irregular de múltiples variables, archivos y medios de forma simultánea, entreviendo patrones en dinámicas ocultas que no son visualmente apreciables.

El uso de técnicas de inteligencia artificial se encuentra en constante expansión en la actualidad, debido a su eficiencia para afrontar problemas complejos en diferentes áreas de estudio, sumado a que las herramientas computacionales modernas permiten una implementación viable de esta técnica. En el campo del análisis forense digital, tal como se evidenció en los resultados del análisis, las técnicas de inteligencia artificial permitieron un entrenamiento de modelos de aprendizaje a partir de la experiencia y la identificación de patrones en los posibles vectores de ataque y alteración de la información, evidenciando las siguientes ventajas:

- Capacidad de realizar tareas con base en criterios adquiridos a partir de un entrenamiento.
- Abstracción de una representación de la información y organización de acuerdo con esta.
- Tolerancia a fallos y composición modular de las técnicas de aprendizaje que permiten que los servicios, a pesar de sufrir daños parciales, tengan capacidades que se puedan conservar.

- Las operaciones y cálculos pueden hacerse en paralelo y tiempo real, lo que permite la identificación en línea de patrones en vectores de ataque.

Por otro lado, se evidencia que la inteligencia artificial es aplicable a múltiples contextos dentro del análisis forense digital, lo que permite reforzar esta disciplina. Desde la identificación de aspectos físicos mediante el procesamiento de imágenes de las escenas, hasta la detección de alteraciones en archivos y datos. Esto demuestra que la inteligencia artificial es un área flexible y heterogénea en técnicas que se pueden aplicar a variados contextos, que está en constante evolución y que ha colaborado con la solución de algunos inconvenientes que se presentan a la hora de realizar tareas humanas.

Sin embargo, aunque las técnicas de inteligencia artificial presentadas resultan eficientes para el análisis y correlación de grandes volúmenes de datos, siguen presentando una tasa de falsos positivos y falsos negativos alta y su implementación implica un alto consumo de recursos computacionales. Adicionalmente, se trata de técnicas basadas en la experiencia, razón por la cual ataques de días cero o vectores de ataque no previstos pueden no ser detectados correctamente al analizar los datos recabados. Por lo tanto, debe hacerse un esfuerzo adicional por seguir implementando la inteligencia artificial de forma integrada con el análisis manual y visual de los datos, pues la base de conocimientos es fundamental para un mejor desempeño de los modelos de detección implementados.

Aunque este estudio presenta reflexiones sobre la problemática planteada, existen algunas limitaciones a destacar: poca literatura que documente el desarrollo y resultados preliminares de pruebas de aplicación de la inteligencia artificial en el ámbito forense digital. Así mismo, se incluyeron técnicas específicas usadas para la solución de problemáticas enmarcadas en el análisis forense digital, obviando el uso de otras técnicas como los modelos de Markov, estadísticas Bayesianas y modelos de redes neuronales más robustos, entre otros. El análisis del desarrollo e implementación de técnicas de inteligencia artificial se hace de forma limitada, sin entrar en detalle en áreas específicas del análisis forense digital, como el almacenamiento en la nube, operaciones transaccionales y bancarias, o procesos cuánticos criptográficos, entre otros.

VI. CONCLUSIONES

En este documento se detallan las características y propiedades de los modelos de inteligencia artificial que son aplicados al análisis forense digital. Se presenta una breve revisión de la literatura para ilustrar en qué áreas del análisis forense digital se ha utilizado recientemente la inteligencia artificial, proporcionando una visión de las múltiples disciplinas de trabajo del aprendizaje automático, las máquinas de soportes vectoriales y las redes neuronales artificiales, así como los agentes inteligentes.

Muchos son los retos que aún debe resolverse con el uso de técnicas automatizadas que permitan la identificación de patrones nacientes, ataques e información extraíble de

dinámicas ocultas no visibles mediante la implementación de métodos manuales.

En este sentido, este breve estudio establece un punto de referencia para habilitar nuevas perspectivas del uso de la inteligencia artificial en el análisis forense digital, que permitan orientar investigaciones posteriores y propuestas novedosas de aplicaciones emergentes, que atendiendo las ventajas y limitaciones de las disciplinas de aprendizaje automático y los agentes inteligentes revisadas, puedan facilitar la convergencia entre los métodos manuales y las posibilidades que plantean las estrategias automáticas revisadas.

AGRADECIMIENTOS

Esta investigación estuvo soportada por la Universidad Pontificia Bolivariana de Bucaramanga. Gracias al Maestro Diego Javier Parada, Coordinador del Programa de Especialización en Seguridad Informática, por el apoyo y colaboración en el desarrollo de esta investigación.

REFERENCIAS

- [1] M. López Delgado, *Análisis Forense Digital*, España: CriptoRed, 2007.
- [2] S. Raghavan, «Digital forensic research: Current state of the art,» *CSI Transactions*, pp. 91-114, 2013.
- [3] M. J. Rivas Sáñez, «A Review of Technical Problems when Conducting an Investigation in Cloud-Based Environments,» *arXiv*, p. 16, 2014.
- [4] H. Ç. a. M. A. S. Dilek, «Applications of Artificial Intelligence Techniques to combating Cyber Crimes: a review,» *International Journal of Artificial Intelligence & Applications (IJAI)*, vol. 6, pp. 21-23, 2015.
- [5] L. D. Merkle, «Automated Network Forensics,» pp. 1929-1931, 2008.
- [6] . A. Kaplan y M. Haenlein, «Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations,» *Illustrations and Implications of Artificial Intelligence, Business Horizons*, vol. 62, n° 1, pp. 15-25, 2019.
- [7] D. Poole, *Computational Intelligence: A Logical Approach*, Nueva York: Oxford University Press, 2018.
- [8] N. Nilsson, *Introduction to Machine Learning*, Stanford, California: Department of Computer Science, Stanford University, 2005.
- [9] P. Harrington, *Machine Learning in Action*, United States of America: Manning publications Co., 2012, p. 382.
- [10] D. Kriesel, *A Brief Introduction to Neural Networks*, Bonn, Germany: Dkriesel, 2005, p. 244.
- [11] A. Mammone, M. Turchi y N. Cristianini, «Support Vector Machines,» vol. 1, pp. 283-288, 2009.
- [12] A. Navia Vazquez y E. Parrado Hernandez, «Support vector machine interpretation,» *Neurocomputing*, pp. 1754-1759, 2006.
- [13] D. Matich, *Redes Neuronales: Conceptos Básicos y Aplicaciones*, Rosario: Universidad Tecnológica Nacional, 2001, p. 55.

- [14] C. A. Ruiz y M. S. Basualdo, *Redes Neuronales: conceptos básicos y aplicaciones*, Rosario: Universidad Tecnológica Nacional, Facultad Regional Rosario, 2001, p. 55.
- [15] C. S. Krishnamoorthy y S. Rajeev, *Artificial Intelligence and Expert Systems for Engineers*, Boca Ratón, Florida: CRC Press, 2000, p. 190.
- [16] L. C. Jain, Z. Chen y N. Ichalkaranje, *Intelligent Agents and Their Applications*, New York: Physica-Verlag, Springer, 2002.
- [17] S. M. a. A. H. Sung, «Identifying Significant Features for Network Forensic Analysis Using Artificial Intelligent Techniques,» *International Journal of Digital Evidence*, vol. 1, n° 4, 2003.
- [18] A. Mikkilineni, O. Arslan, P.-J. Chiang, R. Kumontoy, J. Allebach, G. Chiu y E. Delp, *Purdue University*, p. 4, 2005.
- [19] Q. Li y A. Ong, «A Novel Support Vector Machine Approach to High Entropy Data Fragment Classification,» *Proceedings of the South African Information Security Multi-Conference*, p. 10, 2010.
- [20] K. Cai, X. Lu, J. Song y X. Wang, «Blind Image Tampering Identification Based on Histogram Features,» de *Third International Conference on Multimedia Information Networking and Security*, Shanghai, China, 2011.
- [21] L. Sportiello y S. Zanero, «File Block Classification by Support Vector Machines,» de *Sixth International Conference on Availability, Reliability, and Security*, Milan, Italy, 2011.
- [22] R. Mohammad, «A Neural Network-based Digital Forensics Classification,» *IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, p. 7, 2018.
- [23] A. Tallón-Ballesteros y J. Riquelme, «Data Mining Methods Applied to a Digital Forensics Task for Supervised Machine Learning,» *Computational Intelligence in Digital Forensics: Forensic Investigation and Applications*, vol. 555, pp. 413-428.
- [24] Y. Qian, J. Dong, W. Wang y T. Tan, «Learning and Transferring Representations for Image Steganalysis using Convolutional Neural Network,» de *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016.
- [25] D.-H. Kim y H.-Y. Lee, «Image Manipulation Detection using Convolutional Neural Network,» *International Journal of Applied Engineering Research*, vol. 12, n° 21, pp. 11640-11646, 2017.
- [26] N. Bonettini, «JPEG-based Forensics through Convolutional Neural Networks,» Milano, 2017.
- [27] J. Bunk, J. Bappy, T. Mohammed, L. Nataraj, A. Flenner, B. Manjunath, S. Chandrasekaran, A. Roy-Chowdhury y L. Peterson, «Detection and Localization of Image Forgeries using Resampling Features and Deep Learning,» de *CVPR Workshop on Media Forensics*, University of Maryland, 2017.
- [28] B. Hoelz, C. Ralha, R. Geeverghese y H. Junior, «A Cooperative Multi-Agent Approach to Computer Forensics,» *IEEE International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 1-7, 2008.
- [29] S. Qatawneh, S. Ipson, R. Qahwaji y H. Ugail, «3D face recognition based on machine learning,» *Eighth IASTED International Conference on Visualization, Imaging and Image Processing (VIIP 2008)*, pp. 362-366, 2008.
- [30] D. Saez Trigueros, L. Meng y M. Hartnett, «Face Recognition: From Traditional to Deep Learning Methods,» *Cornell University Press*, p. 13, 2018.
- [31] F. Mitchell, «THE USE OF ARTIFICIAL INTELLIGENCE IN DIGITAL FORENSICS: AN INTRODUCTION,» *Digital Evidence and Electronic Signature Law Review*, vol. 7, pp. 35-41, 2010.
- [32] F. Spencer, «Digital Forensics with Artificial Intelligence Internet of Things,» p. 6, 2018.
- [33] B. Hoelz, R. Geeverghese y C. G. Ralha, «Artificial intelligence applied to computer forensics,» *Proceedings of the 2009 ACM Symposium on Applied Computing (SAC)*, p. 6, 2009.
- [34] J. Li, S. You y A. Robles-Kelly, «A Frequency Domain Neural Network for Fast Image Super-resolution,» *International Joint Conference on Neural Networks*, p. 9, 2017.
- [35] F. Petroski Such, S. Sah, M. Dominguez, S. Pillai, C. Zhang, A. Michael, N. Cahill y R. Ptucha, «Robust Spatial Filtering with Graph Convolutional Neural Networks,» *IEEE Journal of Selected Topics in Signal Processing*, p. 14, 2017.
- [36] L. Özgür, T. Gungor y F. Gergen, «Spam Mail Detection Using Artificial Neural Network and Bayesian Filter,» *Intelligent Data Engineering and Automated Learning - IDEAL, 5th International Conference*, p. 6, 2004.
- [37] S. Sekhar Roy y M. Viswanatham, «Classifying Spam Emails Using Artificial Intelligent Techniques,» *International Journal of Advanced Computer Technology (IJACT)*, p. 5, 2016.
- [38] K. Al Smadi, H. A. Al Issa, I. Trrad y T. Al Smadi, «Artificial Intelligence for Speech Recognition Based on Neural Networks,» *Journal of Signal and Information Processing*, p. 7, 2006.

Jeimy J. Cano. Egresado del Programa de Ingeniería y Maestría en Sistemas y Computación de la Universidad de Los Andes. Doctor en Filosofía de la Administración de Negocios, de Newport University en California, Estados Unidos. Certificado como Examinador Certificado de Fraude - en inglés CFE. Es profesor e investigador a nivel nacional y latinoamericano en temas de seguridad informática, computación forense y sistemas de información. Actualmente, es director de la revista SISTEMAS, de la Asociación Colombiana de Ingenieros de Sistemas (ACIS).

Julián D. Miranda. Ingeniero Electrónico (2016), Ingeniero de Sistemas e Informática (2018) y Especialista en Seguridad Informática (2019) de la Universidad Pontificia Bolivariana de Bucaramanga, Colombia. Cuenta con experiencia en el desarrollo de proyectos de investigación con metodologías Ágiles, aplicando aprendizaje automático y procesamiento digital de imágenes y señales, enfocadas hacia la Inteligencia Artificial y las ciencias de datos. Actualmente, es docente de pregrado y posgrado en las áreas de computación, sistemas operativos, criptografía, esteganografía y ciencia de datos.

Sergio A. Pinzón. Ingeniero de Telecomunicaciones (2013) de la Universidad Santo Tomas de Aquino de Bucaramanga, Colombia. Cuenta con experiencia en Arquitectura de redes móviles, redes cableadas de fibra óptica y cableado UTP, configuración de dispositivos de comunicaciones y seguridad perimetral, con intereses en temas de Machine Learning aplicado a la seguridad de la información, automatización de dispositivos de red y Ethical Hacking.

MARISMA-BiDa: Gestión y Control del riesgo en Big Data. Caso de Estudio

David G. Rosado, Julio Moreno, Luis E. Sánchez, Antonio Santos-Olmo, Manuel A. Serrano, Eduardo Fernández-Medina

Resumen— En la actualidad, se genera una gran cantidad de información debido a la amplia hiperconectividad y sensorización del mundo que nos rodea. Esta información es considerada como uno de los activos más importantes para las empresas en todos los campos. El continuo crecimiento en la importancia y el volumen de datos ha creado un nuevo problema: no puede ser manejado por las técnicas de análisis tradicionales. Este problema se resolvió, por lo tanto, mediante la creación de un nuevo paradigma: Big Data. Sin embargo, Big Data originó nuevos problemas relacionados no sólo con el volumen o la variedad de los datos, sino también con la seguridad y privacidad de los datos. Al adoptar nuevas soluciones tecnológicas como Big Data, todos los riesgos deben ser identificados y gestionados. En este artículo se presenta un caso de estudio de la aplicación de una técnica de análisis y gestión de riesgos para entornos Big Data, guiada por una metodología de gestión de la seguridad (MARISMA) y soportada por un entorno tecnológico en la nube (eMARISMA). La propuesta, denominada MARISMA-BiDa es un patrón específico para Big Data que contiene los elementos necesarios para facilitar la aplicación de la metodología de análisis y gestión de riesgos MARISMA en un entorno específico y siguiendo los principales estándares y recomendaciones internacionales relacionados con Big Data (ISO/IEC, NIST, ENISA).

Palabras clave—Big Data, Análisis y Gestión de riesgos, Seguridad

I. INTRODUCCIÓN

UN número creciente de dispositivos, sensores y personas están conectados a la red global y esto cambia drásticamente la capacidad de generar, comunicar, compartir y acceder a los datos [1]. Los datos son esenciales para el desarrollo de sus actividades cotidianas, así como para ayudar a la dirección de las empresas a alcanzar sus objetivos y a tomar las mejores decisiones a partir de la información que se extrae de ellas [2]. Los desarrollos tecnológicos y nuevas aplicaciones continúan alimentando el debate sobre lo que define a Big Data y lo distingue de las anteriores formas de análisis de datos [3]. No existe un consenso real en cuanto a

sus características clave, aunque la mayoría de las definiciones de Big Data se refieren a las tres Vs [4]. La primera V es para Volumen (el uso de grandes cantidades de datos), la segunda V es para Variety (el uso de diversas fuentes de datos que se almacenan en estructuras diversas o incluso de forma no estructurada) y la tercera significa Velocidad, o la velocidad del procesamiento de datos (los datos son a menudo en tiempo real). A lo largo del tiempo, varios autores han añadieron Vs adicionales a este trío, como Veracity [5], Variabilidad [6], [7], Valor [8], [9] y Virtual [10], [11].

La aplicación de Big Data ofrece beneficios significativos para los individuos y la sociedad, pero también plantea serias preocupaciones sobre varios riesgos de seguridad de la información como la seguridad de los datos, el gobierno y la privacidad [12]. Uno de los principales problemas en el uso de los sistemas Big Data es la seguridad. Los sistemas Big Data son complejos y heterogéneos, y la seguridad de todo el sistema debe ser abordada de manera integral. Además, la integración de diferentes tecnologías introduce nuevas cuestiones de seguridad que deben abordarse adecuadamente [1]. Big Data no fue diseñado con la seguridad en mente. Con estas montañas de datos, que informan a las empresas sobre las decisiones críticas de los clientes, los hábitos y otros innumerables detalles, surge la necesidad urgente de mantener esta valiosa información segura y protegida. Al fin y al cabo, se trata de información delicada, y con gran parte de ella existe un mayor riesgo de infracciones [13].

Los problemas de seguridad y privacidad se ven magnificados por la velocidad, el volumen y la variedad de los grandes datos, como las infraestructuras de nube a gran escala, la diversidad de fuentes y formatos de datos, la naturaleza de la transmisión de la adquisición de datos y la migración entre nubes de gran volumen. El uso de infraestructuras de nube a gran escala, con una diversidad de plataformas de software, distribuidas en grandes redes de ordenadores, también aumenta la superficie de ataque de todo el sistema. Por tanto, es muy importante contar con una serie de guías, metodologías y mecanismos para implementar de forma adecuada no solo el entorno Big Data, sino también su seguridad. Pero no solo eso, además, es ampliamente considerado que todo entorno global de gestión de seguridad de la información en la empresa, debe estar centrado en los riesgos [14]–[16]. Por lo que los riesgos de seguridad en Big Data, deben ser analizados y gestionados de manera adecuada, junto a los riesgos de otros tipos de activos de información [17].

Por otro lado, la mayoría de las organizaciones de hoy en

David G. Rosado, Julio Moreno, Luis E. Sánchez y Eduardo Fernández-Medina, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, david.grosado@uclm.es, julio.moreno@uclm.es, luisenrique@sanchezcrespo.org, eduardo.fdezmedina@uclm.es.

Antonio Santos-Olmo, Departamento I+D+i, Sicaman Nuevas Tecnologías y Marisma Shield, Tomelloso (Ciudad Real), España, asolmo@sicaman-nt.com.

Manuel A. Serrano, Grupo de Investigación Alarcos, Universidad de Castilla-la Mancha, Ciudad Real, España, Manuel.serrano@uclm.es.

día que utilizan tecnologías de la información tienen problemas con la seguridad de su sistema de información, y diferentes investigadores destacan que la gestión del riesgo es un proceso esencial en cualquier modelo de gestión empresarial [18], y que la información es un activo valioso que se espera que esté protegido [19].

Un análisis de riesgos es un proceso sistemático para estimar la magnitud de los riesgos a los que está expuesta una organización, para saber qué decisión tomar ante una posible eventualidad [20]. Para ello, se seleccionan e implementan salvaguardas para poder conocer, prevenir, impedir, reducir o controlar los riesgos identificados. Esto es lo que se entiende como gestión de riesgos.

De forma más técnica, el análisis de riesgos permite determinar cómo es, cuánto vale y cómo de protegidos se encuentran los activos. En coordinación con los objetivos, estrategia y política de la organización, las actividades de gestión de riesgos permiten elaborar un plan de seguridad que, implantado y operado, satisfaga los objetivos propuestos con el nivel de riesgo que acepta la dirección.

Actualmente se están realizando muchas investigaciones sobre análisis de riesgos, y muchas de ellas intentan comparar los métodos clásicos para ver cómo se podrían alinear [21]–[27]. Otros investigadores han realizado también algunos análisis comparativos de los principales estándares de riesgos con el objetivo de mejorar algunos de sus aspectos [28], o trabajos que relacionan los planes de contingencia con el análisis de riesgos [29]. Durante el estudio, se identificaron ciertas deficiencias como pueden ser dificultades para su aplicación en la práctica, no cuentan con herramientas adecuadas para su procesamiento (o en caso de existir, éstas no son muy usables), están pensadas para ser aplicadas en grandes compañías, no son sensibles al contexto, sin contar con capacidades de adaptación para entornos especiales, que requieran un especial tratamiento de sus riesgos, y su falta de dinamismo y asociatividad de sus riesgos. Parte de estas deficiencias se afrontaron mediante el desarrollo de una metodología denominada MARISMA (Methodology for the Analysis of Risks on Information System, using Meta-Pattern and Adaptability) que utilizan el concepto de meta-patrón para crear una estructura capaz de soportar los elementos asociados con un análisis de riesgos y sus relaciones, con el objetivo de crear estructuras evolutivas, dinámicas y capaces de adaptarse a las nuevas tecnologías [30], [31].

Dentro de estas nuevas tecnologías ha tomado especial importancia la necesidad de poder analizar los riesgos de seguridad dentro de los entornos de Big-Data, y por tanto se ha considerado relevante, afrontar el desarrollo de un nuevo patrón soportado sobre la metodología MARISMA, que permita realizar análisis de riesgos TIC en entornos de Big-Data, validándolo en un caso de estudio relacionado con el sector sanitario, donde los datos que se manejan son muy sensibles y el nivel de seguridad que se requiere es muy alto [32], [33].

El resto de este artículo está estructurado de la siguiente manera: La sección II presenta el marco de trabajo MARISMA donde se define el modelo conceptual de gestión

de riesgos y se explica la metodología de análisis y gestión de riesgos utilizada. Además, se describe el meta-patrón definido y la herramienta de soporte utilizada. La sección III se define el patrón MARISMA-BiDa usando el meta-patrón para el contexto de Big Data definiendo los elementos específicos en estos entornos. En la sección IV se presenta un caso de estudio de registros médicos utilizando el patrón MARISMA-BiDa mostrando los resultados obtenidos y su implementación en la herramienta de soporte. Finalmente, la sección V muestra las conclusiones de nuestra investigación.

II. MARCO DE TRABAJO MARISMA

En los últimos años se han detectado una serie de deficiencias en las principales propuestas de procesos y métodos de gestión de riesgos. Entre los principales problemas identificados en los métodos de Gestión de Riesgos, hay algunos que se pueden destacar: i) Alto costo y complejidad, ya que se realiza el análisis de riesgos, ii) Su falta de orientación hacia las pequeñas y medianas empresas (PYMES); iii) Los resultados de los conocimientos de gestión de riesgos de proyectos anteriores no suelen considerarse más fáciles para ejecutar nuevos procesos de gestión de riesgos, y iv) Los análisis de riesgos son casi siempre estáticos. En consecuencia, estas cuestiones ponían en tela de juicio su eficacia y ponían en peligro su valor para las organizaciones.

Para resolver estos problemas, se desarrolló una metodología llamada MARISMA, así como una herramienta de apoyo a dicha metodología (herramienta eMARISMA¹). Esta metodología se centra en la reducción de los costes del proceso de gestión de riesgos y en la simplicidad de su aplicación; el modelo desarrollado permite el mayor nivel posible de automatización y reutilización con la mínima cantidad de información, recogida en un tiempo muy reducido [34].

En las siguientes subsecciones se muestran los detalles sobre esta metodología y esta herramienta.

A. Modelo Conceptual de Gestión de Riesgos

Las organizaciones, independientemente de su tamaño, deben ser conscientes de la importancia de los riesgos de TI y de cómo deben gestionarse. La ISO 31000 [35] define la gestión de riesgos como un proceso organizacional que debe implicar la aplicación sistemática de políticas, procedimientos y prácticas a las actividades de comunicación y consultoría, estableciendo el contexto y evaluando, tratando, monitoreando y revisando el riesgo.

Todas estas actividades deben ser controladas por los actores internos y externos del proceso de gestión de riesgos [35], [36]. La Fig. 1. resume este proceso incluyendo todas las actividades y sus relaciones.

¹ www.emarisma.com

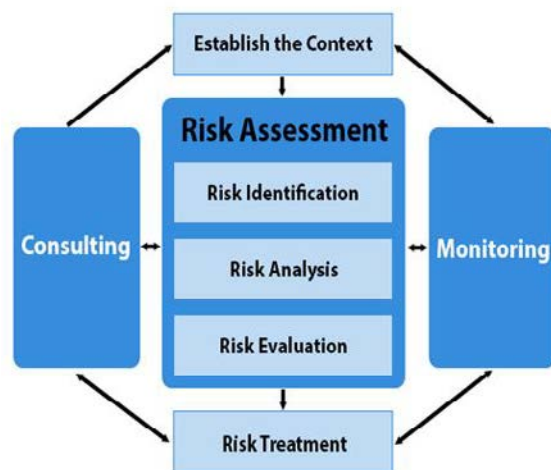


Fig. 1 Proceso de Gestión de Riesgos basado en ISO 31000

Siguiendo el proceso de gestión de riesgos anteriormente expuesto, y teniendo en cuenta todas las actividades, se ha considerado el desarrollo de MARISMA, ya que se trata de un marco que da soporte a todo el proceso de gestión de riesgos, centrándose principalmente en la actividad de tratamiento de riesgos. Se propone un modelo conceptual de gestión de riesgos en la Fig. 2, que representa los conceptos esenciales que deben ser capturados -en forma de información- para

automatizar el proceso de análisis de riesgos a través de la herramienta eMARISMA. Este modelo conceptual sirve para establecer las relaciones entre todos los conceptos, y es posible almacenar toda la información que se genera en cualquier proceso de análisis de riesgos.

Como se muestra en la Fig. 1, el contexto de la organización es crucial para definir adecuadamente sus activos. Estos son los elementos más importantes de la gestión de riesgos y son esenciales para protegerlos. Conociendo los activos involucrados en la organización, se pueden identificar varias vulnerabilidades asociadas con ellos. Estas vulnerabilidades pueden ser mitigadas con el apoyo de controles de seguridad. Sin embargo, para abordar adecuadamente estas vulnerabilidades, es aconsejable abstraer los activos en un concepto más general, a saber, los tipos de activos. Los diferentes tipos de activos pueden estar en riesgo debido a amenazas específicas (clasificadas por categorías o tipos de amenazas), pero este riesgo no compromete a activos particulares en su totalidad, sino que lo hace en diferentes dimensiones (es decir, confidencialidad, integridad, etc.). Estas dimensiones dependen del contexto y se definen en la subsección 4.3. Es imperativo no olvidar la importancia de los requisitos de seguridad en la identificación y definición de los diferentes controles que tratan de mitigar las amenazas identificadas. El modelo definido se muestra en la Fig. 2 y se ha implementado en el marco de MARISMA.

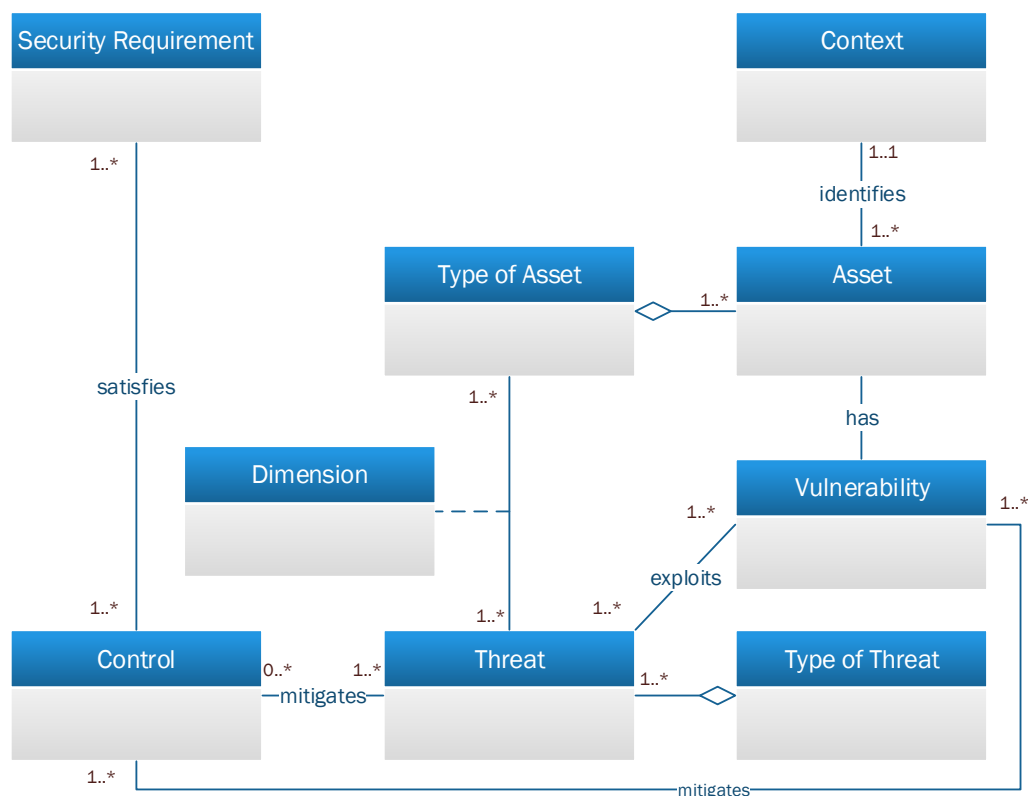
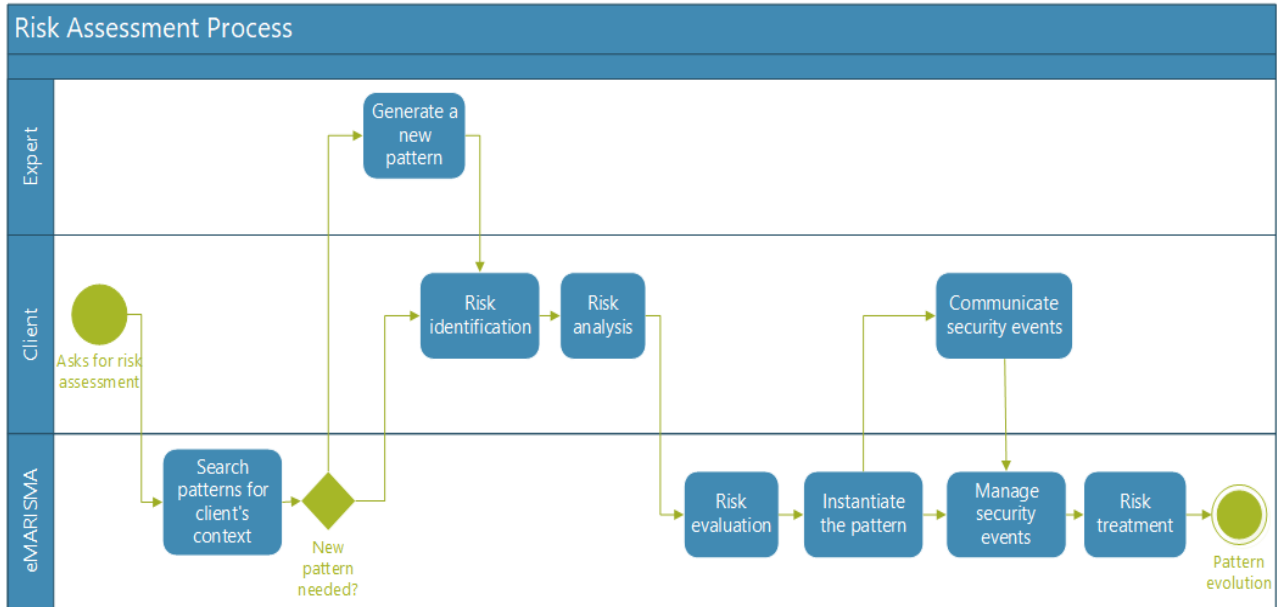


Fig. 2. Modelo conceptual de Gestión de Riesgos



B. Metodología MARISMA

Dado que la gestión de riesgos puede considerarse como

C. Meta-Patrón MARISMA

Fig. 3. Visión general de los procesos en MARISMA

parte de la gestión de la seguridad, MARISMA intenta asociar las actividades de ambos procesos. Para lograr este objetivo se define el siguiente proceso en la Fig. 3. El éxito del proceso se basa en la reutilización de patrones, que son estructuras de conocimiento con características comunes para un contexto específico, por ejemplo, un patrón de riesgo para Big Data, un patrón de riesgo para sistemas críticos, o un patrón de riesgo para servicios web, etc.

El proceso tiene tres puntos de vista: El punto de vista del *Cliente* define todas las iteraciones y el trabajo que el cliente debe realizar durante la identificación y análisis de los riesgos del sistema (identificación de activos, amenazas, controles, incluyendo probabilidad e impacto, etc.). Esto se apoya en la herramienta eMARISMA que ayuda al cliente a calcular el nivel de riesgo y su evaluación para la toma de decisiones. El punto de vista de eMARISMA se encarga de gestionar automáticamente toda la información proporcionada por el cliente a través de patrones de conocimiento, gestionar todos los eventos de seguridad que pueden ocurrir en el sistema, y hacer recomendaciones sobre la mejor forma de tratarlos. El tercer punto de vista es el del *Experto*, cuya única misión en el proceso es generar un nuevo patrón inexistente que contenga las características comunes de un entorno empresarial similar, como un entorno Big Data.

A continuación, el sistema está preparado para gestionar el evento de seguridad que debe ser comunicado por el cliente. Estos eventos de seguridad generan conocimiento para adaptar los niveles asociados a los elementos de la gestión de riesgos, haciendo que el riesgo sea recalculado dinámicamente, y también adaptando los elementos asociados al patrón seleccionado permitiendo su evolución.

Como se definió en [17], el meta-patrón MARISMA está formado por los elementos comunes y sus relaciones que cualquier análisis de riesgos debe tener, con el objetivo de reutilizar el conocimiento y la experiencia adquirida por los consultores al ejecutar el proceso de análisis de riesgos. Así se pudo identificar que todos los análisis de riesgos tenían controles, activos de información y amenazas, y que estos elementos estaban interrelacionados, como se muestra en nuestro modelo conceptual de la Fig. 2.

Por tanto, para diferenciar entre un patrón y un meta-patrón en nuestro marco de trabajo MARISMA, un patrón contiene, por lo tanto, los elementos necesarios para llevar a cabo un proceso de análisis y gestión de riesgos en un contexto específico. Un patrón se basa en un meta-patrón más general que contiene todos los elementos necesarios y sus relaciones para un análisis de riesgos sin tener en cuenta el contexto específico de cada organización o empresa. La organización o empresa es la encargada de crear el patrón específico que mejor se adapte al contexto (punto de vista del experto en la Fig. 3), instanciando el meta-patrón y tomando los elementos que considere necesarios para llevar a cabo el análisis de riesgos de su empresa u organización. Este meta-patrón se muestra en la Fig. 4 y ha sido implementado en la herramienta eMARISMA.

La definición de cada uno de los elementos que forman el meta-patrón ha sido descrito en detalle en [17], y se ha definido siguiendo nuestro modelo conceptual (mostrado en la) que define todos los conceptos para realizar una gestión de riesgos. De esta forma, el meta-patrón contiene los elementos necesarios para representar los conceptos esenciales y sus relaciones en un análisis de riesgos.

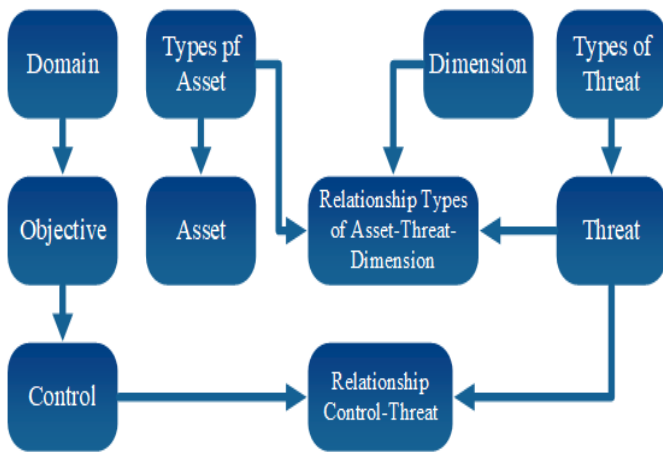


Fig. 4 Meta-patrón en MARISMA

Los elementos que aparecen definidos en la Fig. 4 tienen una relación directa con el modelo conceptual de la de la siguiente forma:

- Dominio y Objetivo de Control: Estos elementos se definen a partir del concepto de *Context* de nuestro modelo conceptual.
- Control: En nuestro modelo conceptual hemos definido el concepto de *Control*.
- Tipos de Activos y Dimensiones: En el modelo conceptual, hay tres conceptos como *Asset*, *Types of Asset* y *Dimension* que los representan.
- Tipos de Amenazas y Amenazas: Estos conceptos están definidos en nuestro modelo conceptual como *Threat* y *Types of Threat*.
- La relación Control-Amenaza: Con este elemento podemos gestionar el concepto de *Vulnerability* para nuestro modelo conceptual.
- La relación Tipos de Activos - Amenaza - Dimensión: Este elemento representa la relación entre *Types of Asset*, *Threat* y *Dimension* de nuestro modelo conceptual.

Estos elementos son necesarios y suficientes para iniciar el proceso de análisis de riesgos y pueden definir el resto de las relaciones de nuestro modelo conceptual como puede ser "mitiga" entre los conceptos de Control y Vulnerabilidad, y la relación "tiene" entre Activo y Vulnerabilidad. El concepto de "Vulnerabilidad" se obtendrá a través de las relaciones existentes del meta-patrón conociendo la falta de controles y amenazas por parte del elemento relación *control-amenaza*.

D. Herramienta eMARISMA

Para apoyar la metodología MARISMA, se desarrolló una herramienta llamada eMARISMA como SaaS y utilizando tecnología Java, que soporta todos los procesos de la metodología, con bajo coste de mantenimiento, dinámico y monitorización en tiempo real.

eMARISMA dispone de diferentes zonas donde el usuario puede gestionar toda la información relacionada con la gestión de riesgos, la visualización y creación de patrones, la generación de relaciones entre conceptos, y el cálculo del

riesgo y los eventos de seguridad. También dispone de una zona de visualización y cuadro de mando. En el caso de estudio podemos ver las capturas de pantalla de la herramienta. Las principales funciones de la herramienta eMARISMA son:

- Visualizar los diferentes patrones existentes y utilizarlos como base para crear otros patrones (por ejemplo, patrones sectoriales).
- Obtener un mapa detallado de la situación actual (identificación de riesgos) y un plan de recomendaciones sobre cómo mejorarla (evaluación de riesgos).
- El sistema realiza automáticamente una evaluación de riesgos y calcula el plan de tratamiento de riesgos más adecuado para que la empresa alcance un nivel de riesgo dentro de los límites definidos de una forma óptima.
- La herramienta representa un cuadro de mando con los niveles de seguridad que la empresa tiene en todo momento, para que el riesgo de la empresa pueda ser monitorizado en tiempo real.

III. PATRÓN MARISMA-BiDA

La versatilidad de la metodología MARISMA permite el desarrollo de un marco de gestión de seguridad especializado en el contexto de Big Data. Para ello, se definiría un patrón especializado (llamado MARISMA-BiDa). Este es un marco genérico que puede aplicarse a cualquier ecosistema de Big Data.

Como hemos comentado anteriormente, el meta-patrón definido es genérico y válido para cualquier contexto, y debe ser el experto el que cree un patrón dependiente del contexto basado en este meta-patrón e instanciarlo con los conceptos específicos relacionados con la empresa. Así, por ejemplo, para un contexto de Big Data, el experto debe crear un patrón con dominios, objetivos y controles específicos, un conjunto de amenazas para Big Data, así como los activos más comunes en este tipo de entornos y sus dimensiones (por ejemplo, Volumen y Velocidad). El cliente utilizará este nuevo patrón e instanciará estos elementos con nombres específicos, por ejemplo, MongoDB como activo, robo de identidad como amenaza, etc.

Todos estos elementos instanciados han sido definidos en un trabajo previo [17] donde por la parte de los dominios, objetivos y controles se ha tomado como base la normativa ISO/IEC 27000; para las dimensiones se han considerado las diferentes Vs típicas de los sistemas Big Data [37]–[39]; para especificar los diferentes tipos de activos se han considerado principalmente la arquitectura de referencia para Big Data realizada por la organización NIST [40] junto con ENISA [41] y Cloud Security Alliance (CSA) [42]; para definir los diferentes tipos de amenazas que pueden afectar a un entorno Big Data se han seguido las recomendaciones dadas por la Agencia de Seguridad de las Redes y de la Información de la Unión Europea (ENISA) [43].

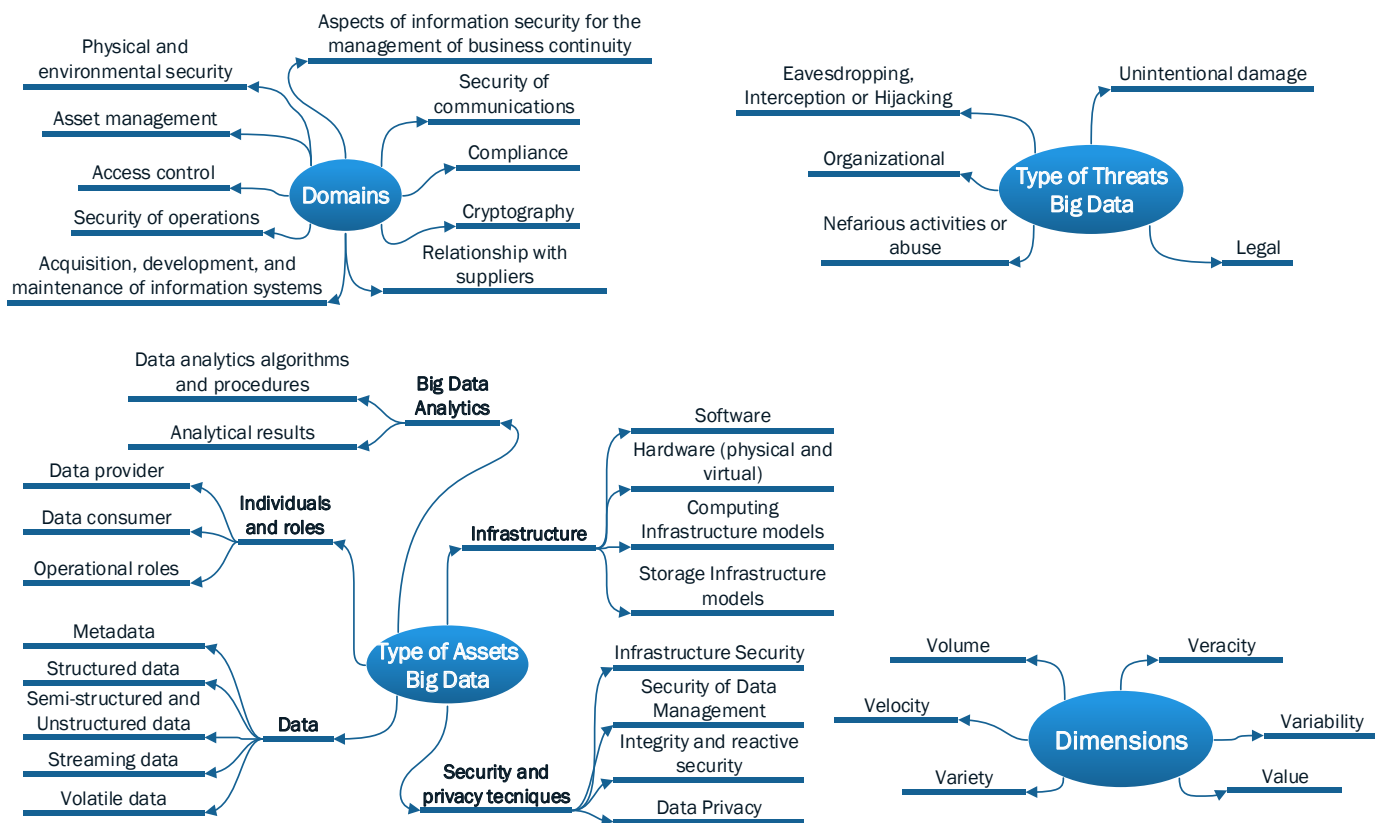


Fig. 5 Elementos que forman parte del patrón MARISMA-BiDa

La Fig. 5 muestra todos los elementos que forman parte del patrón de forma gráfica. Así se pueden ver el conjunto de dominios, los tipos de activos categorizados por grupos de activos, los tipos de amenazas y las dimensiones a tener en cuenta en el patrón para Big Data. Los detalles de cada uno de estos elementos y subelementos se pueden obtener en [17].

Para la definición de las diferentes amenazas que se pueden identificar en este tipo de sistemas, se ha definido una matriz en la que se integran tanto los diferentes tipos de amenazas como las diferentes dimensiones de Big Data. Si una amenaza no encaja en ninguna de las celdas de la matriz, se puede concluir que no es una amenaza típica de un entorno Big Data.

La Tabla 1 muestra un ejemplo de cómo se pueden identificar las diferentes amenazas para los entornos de Big Data relacionadas con una o varias dimensiones. Esta relación se ha definido mediante la realización de un amplio estudio tanto de las amenazas como de las dimensiones sin perder de vista el contexto y los activos más afectados como consecuencia de la materialización de estas amenazas. Con la experiencia del equipo de investigación en estos temas, junto con la colaboración de expertos en seguridad y con la ayuda de los informes de ENISA, ha sido posible completar satisfactoriamente esta relación.

Por ejemplo, un tipo de amenaza es la "intercepción de información" de la "escucha clandestina", donde puede producirse la interceptación de información en las comunicaciones entre aplicaciones Big Data, y donde los protocolos de comunicación son raramente seguros entre este tipo de aplicaciones (sin el uso de TLS y SSL). Se considera que este tipo de amenaza afecta en mayor medida a las dimensiones "Veracidad" y "Valor". Consideramos que afecta a la dimensión de la "veracidad" porque al interceptar la información en la red (por ejemplo, ha sido modificada) no es posible estar seguro de que esa información sea exacta, lo que puede llevar a tomar una decisión equivocada. También afecta a la dimensión "valor" porque si alguien ha sido capaz de interceptar la comunicación, puede haber revelado la información oculta de los datos y, en consecuencia, el valor se pierde. Este tipo de estudio y razonamiento para cada uno de los tipos de amenazas y amenazas definidas para nuestro patrón MARISMA-BiDa ha sido realizado, cuyos resultados se muestran en la Tabla 1.

Este patrón así definido servirá para ser instanciado en un nuevo proceso de análisis de riesgos con la herramienta eMARISMA en un entorno Big Data como es el caso de estudio que se presenta a continuación.

	Unintentional damage	Eavesdropping	Nefarious	Legal	Organizational
Volume	Configuration error Destruction of records		Abuse of Information Manipulation of hardware and software	Violation of laws or legislation	Skill shortage
Variety	Destruction of records Inadequate design		Misuse of audit tools		Skill shortage
Velocity	Inadequate design Erroneous use	Network Reconnaissance Replay of messages	DoS/DDoS Malicious code Remote activity Manipulation of hardware and software Misuse of audit tools Brute force Generation and use of rogue certificates		Skill shortage
Variability	Unintentional change of data Inadequate design				Skill shortage
Veracity	Leaks of data via Web applications Loss of sensitive information Loss of information in the cloud Damage caused by a third party Using information from an unreliable source	Intercepting compromising emissions War driving Interfering radiation Man-in-the-middle Interception of information	Code injection Social Engineering Abuse of authorizations Receive of unsolicited E-mail Identity theft Hoax Compromising confidential information Generation and use of rogue certificates Unauthorized activities	Failure to meet contractual requirements Unauthorized use of IPR protected resources	Skill shortage
Value	Information leakage Loss of devices, storage media and documents Unintentional change of data Damages resulting from penetration testing Inadequate design	Man-in-the-middle Interception of information	Malicious code Abuse of Information Failed of business process Targeted attacks Manipulation of information Misuse of information Unauthorized installation of software/Manipulation of Algorithms	Judiciary decisions Abuse of personal data	Skill shortage
Affected Assets	• Data • Infrastructure • Big Data analytics	• Data	• Infrastructure • Data • Big Data analytics • Security and privacy techniques	• Data	• Individuals and Roles

Tabla 1. Matriz de Identificación de “Tipos de Activos – Amenazas - Dimensiones” para entornos Big Data.

IV. CASO DE ESTUDIO: DATOS HISTORIA CLÍNICA ELECTRÓNICA

NIST define diferentes casos de uso para un amplio conjunto de dominios de aplicación, y uno de ellos ha sido seleccionado por su interesante aplicación en el patrón MARISMA-BiDa a un caso particular en un entorno Big Data. El caso de estudio seleccionado se refiere a los datos de registros médicos electrónicos cuyo objetivo es “Utilizar métodos avanzados para normalizar la identificación de pacientes, proveedores, instalaciones y conceptos clínicos dentro y entre organizaciones de atención de la salud separadas para mejorar los modelos de definición y extracción de fenotipos clínicos a partir de datos clínicos de

texto libre y discreto no estándar mediante la selección de características, la recuperación de información y los modelos de toma de decisiones de aprendizaje automático. Aprovechar los datos del fenotipo clínico para apoyar la selección de cohortes, la investigación de resultados clínicos y el apoyo a la toma de decisiones clínicas” (Caso de Uso 16 de [44]).

Este caso de estudio influye en todas las dimensiones definidas por el patrón MARISMA-BiDa (como se muestra en la Fig. 6), por ejemplo:

- *Volumen (DIM1)*: más de 12 millones de pacientes, más de 4.000 millones de observaciones clínicas y más de 20 TB de datos brutos.
- *Velocidad (DIM2)*: entre 500.000 y 1,5 millones de nuevas transacciones clínicas en tiempo real añadidas al

día.

- *Variedad (DIM3)*: una amplia variedad de conjuntos de datos clínicos de múltiples fuentes: notas del proveedor de texto libre; informes de laboratorios y servicios de urgencias; estudios de química, cardiología o hematología; estudios de bancos de sangre y toxicología, etc.
- *Veracidad (DIM4)*: los datos de cada fuente clínica se recopilan comúnmente utilizando diferentes métodos y representaciones, lo que produce una heterogeneidad sustancial. Esto conduce a errores y sesgos sistemáticos que requieren métodos robustos para crear la interoperabilidad semántica.
- *Variabilidad (DIM5)*: los datos de los sistemas clínicos evolucionan con el tiempo porque el espacio conceptual clínico y biológico está en constante evolución. Los nuevos descubrimientos científicos conducen a nuevas entidades de la enfermedad, nuevas modalidades de diagnóstico y nuevos enfoques de manejo de la enfermedad.
- *Valor (DIM6)*: Métodos de recuperación de información para identificar características clínicas relevantes, es decir, modelos de decisión utilizados para identificar una variedad de fenotipos clínicos como la diabetes, la insuficiencia cardíaca congestiva y el cáncer de páncreas.

Code	Name	Date Created	Actions
DIM1	Volume	4/2/2019 10:54:22	[New] [Delete]
DIM2	Velocity	4/2/2019 10:54:55	[New] [Delete]
DIM3	Variety	4/2/2019 10:55:10	[New] [Delete]
DIM4	Veracity	4/2/2019 10:55:21	[New] [Delete]
DIM5	Variability	4/2/2019 10:55:40	[New] [Delete]
DIM6	Value	4/2/2019 10:55:50	[New] [Delete]

Fig. 6 Dimensions defined on the eMARISMA tool

Los pasos del proceso MARISMA comienzan con la identificación de los activos involucrados en el sistema del tipo de activos definidos por el patrón MARISMA-BiDa. Este conjunto de activos involucrados se muestra en la Tabla 2, y en la Fig. 7 se muestran los activos añadidos a la herramienta eMARISMA.

Una vez que los activos para nuestro caso de estudio han sido identificados y añadidos a la herramienta eMARISMA, se realiza automáticamente la evaluación de riesgos utilizando todas las relaciones y matrices definidas en nuestro patrón MARISMA-BiDa. La herramienta relaciona entonces el activo definido en nuestro caso con las amenazas y dimensiones afectadas. También muestra el resultado del riesgo actual para este conjunto de activos y vincula los objetivos con los controles necesarios para proteger este conjunto de activos.

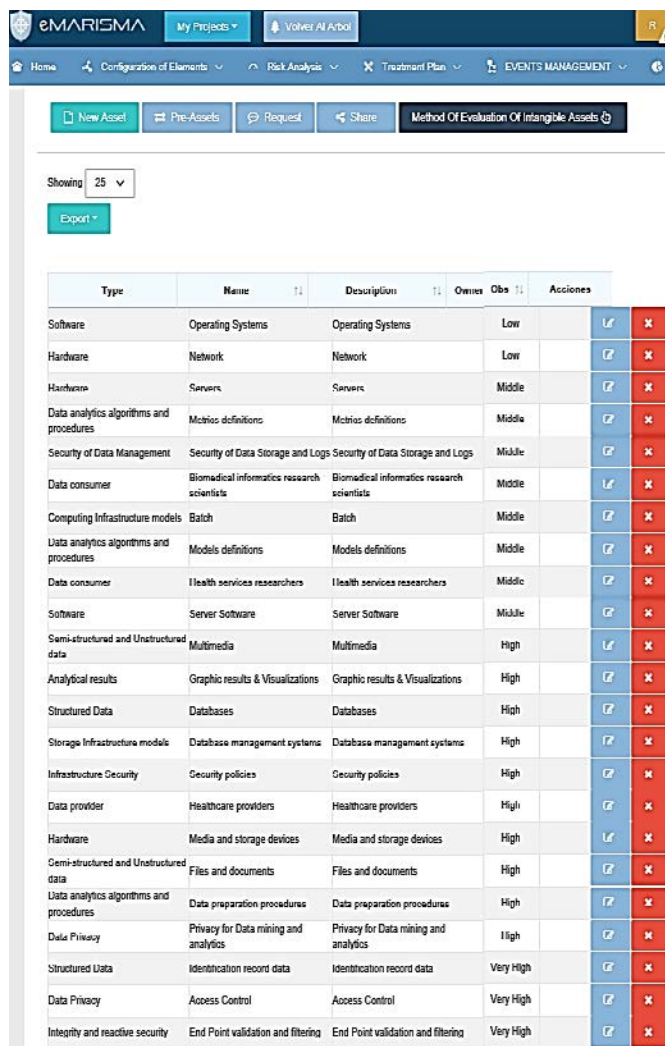
Para el conjunto de amenazas identificadas en el patrón MARISMA-BiDa, la herramienta eMARISMA toma valores por defecto basados en la experiencia de análisis de riesgo previo, así como para la probabilidad de ocurrencia y el porcentaje de degradación, los cuales pueden ser modificados de acuerdo a nuestros criterios, experiencia y entorno del sistema a evaluar.

MARISMA-BiDa PATTERN	PATTERN INSTANTIATED	
Asset group	Asset Type	Assets of Case use
Infrastructure	Software	Operating Systems, Server Software
	Hardware (physical/virtual)	Servers, Network, Media and storage devices
	Computing Infrastructure models	Batch
	Storage Infrastructure models	Database management systems (Teradata, PostgreSQL, MongoDB)
Data	Structured data	Identification record data, Databases
	Semi-structured and Unstructured data	Files and documents Multimedia
Big Data Analytics	Data analytics algorithms and procedures	Metrics definitions, Models definitions, Data preparation procedures
	Analytical results	Graphic results & Visualizations
Security and privacy techniques	Infrastructure Security	Security policies
	Security of Data Management	Security of Data Storage and Logs
	Integrity and reactive security	End Point validation and filtering
	Data Privacy	Privacy for Data mining and analytics, Access Control
Individuals and roles	Data provider	Healthcare providers (physicians, nurses, public health officials)
	Data consumer	Biomedical informatics research scientists, Health services researchers

Tabla 2. Activos involucrados en el caso de estudio para los tipos de activos definidos en el patrón MARISMA-BiDa

La herramienta eMARISMA, con los datos introducidos en el patrón, inicializa todos los valores a valores por defecto, que pueden cambiar con el tiempo a medida que se producen eventos de seguridad, ya que, como se menciona en la sección II.B, la evaluación de riesgos puede recalcularse a medida que se producen eventos de seguridad, como un ataque específico

al sistema. La herramienta muestra los valores por defecto tanto para la probabilidad de ocurrencia de las amenazas como para el porcentaje en que degradan los criterios de riesgo de los activos. Como la herramienta es utilizada por usuarios expertos en seguridad, tienen una idea clara de qué tipos de amenazas son las más comunes en su entorno y cómo puede degradar los activos si la amenaza tiene éxito. De esta manera, pueden cambiar libremente esos valores por otros más acordes con el contexto en el que opera el sistema a discreción de sus expertos. Sin embargo, en una primera evaluación no es necesario cambiar ningún valor, ya que la herramienta realiza una evaluación preliminar con los valores por defecto y los modifica cuando se producen ataques o intentos de ataque.



Type	Name	Description	Owner	Obs	Acciones
Software	Operating Systems	Operating Systems	Low		
Hardware	Network	Network	Low		
Hardware	Servers	Servers	Middle		
Data analytics algorithms and procedures	Metrics definitions	Metrics definitions	Middle		
Security of Data Management	Security of Data Storage and Logs	Security of Data Storage and Logs	Middle		
Data consumer	Biomedical informatics research scientists	Biomedical informatics research scientists	Middle		
Computing Infrastructure models	Batch	Batch	Middle		
Data analytics algorithms and procedures	Models definitions	Models definitions	Middle		
Data consumer	IHealth services researchers	IHealth services researchers	Middle		
Software	Server Software	Server Software	Middle		
Semi-structured and Unstructured data	Multimedia	Multimedia	High		
Analytical results	Graphic results & Visualizations	Graphic results & Visualizations	High		
Structured Data	Databases	Databases	High		
Storage Infrastructure models	Database management systems	Database management systems	High		
Infrastructure Security	Security policies	Security policies	High		
Data provider	Healthcare providers	Healthcare providers	High		
Hardware	Media and storage devices	Media and storage devices	High		
Semi-structured and Unstructured data	Files and documents	Files and documents	High		
Data analytics algorithms and procedures	Data preparation procedures	Data preparation procedures	High		
Data Privacy	Privacy for Data mining and analytics	Privacy for Data mining and analytics	High		
Structured Data	Identification record data	Identification record data	Very High		
Data Privacy	Access Control	Access Control	Very High		
Integrity and reactive security	End Point validation and filtering	End Point validation and filtering	Very High		

Fig. 7 Assets for the case use added on the eMARISMA tool

Por ejemplo, como muestra la Fig. 8, la amenaza “daños causados por terceros” puede actualizar la probabilidad de ocurrencia y el porcentaje de degradación hasta en un 60%. Este porcentaje se obtiene a partir del conocimiento del contexto en el que se encuentra el sistema. En este caso, es de suma importancia considerar que se trata de un contexto en el que muchas personas están involucradas y desean tener acceso a este tipo de datos sensibles, o incluso empresas externas que se encargan de gestionar y proteger este tipo de datos sensibles. Esto puede causar daños involuntarios debido al mal uso de dichos datos al no seguir ninguna política de seguridad, o a una mala configuración en la base de datos que los almacena, o simplemente debido a la eliminación accidental de tales datos que causaría un daño significativo con respecto a los datos médicos, incluyendo una violación de la ley, ya que este tipo de datos sensibles no están bien protegidos. Por lo tanto, el usuario experto puede considerar que el porcentaje de degradación es alto (60%) y que la probabilidad de que ocurra es mayor de lo esperado, por ejemplo, 60% también. Estos valores se modifican en la herramienta porque se consideran apropiados para el entorno, y el nivel de riesgo se evalúa con los valores introducidos, considerando todos los elementos definidos para nuestro patrón (activos, dimensiones, controles, etc.).

El usuario experto puede modificar o actualizar los valores necesarios debido a las particularidades del escenario considerado y por el amplio conocimiento del sistema y del entorno.

Una vez seleccionados los valores más adecuados para las amenazas, la herramienta eMARISMA aplica la matriz “Tipo de Activos-Amenazas-Dimensiones” para calcular el porcentaje de degradación pero, esta vez, incluyendo las dimensiones afectadas para cada amenaza y tipo de activo definido en el patrón MARISMA-BiDa que se puede ver en la Tabla 3. Como se ha mencionado anteriormente, estos valores pueden ser modificados para cada dimensión de acuerdo a nuestros criterios, experiencia y conocimiento del entorno. La Fig. 9 muestra la matriz con los valores de la herramienta eMARISMA.

Finalmente, la Fig. 10 y la Fig. 11 muestran los informes creados por la herramienta. Representa diagramas kiviart y mapa de riesgos para los niveles de seguridad que la organización de salud tiene en cada momento, por lo que puede ser considerado como un tablero de control para monitorear el riesgo de la organización en tiempo real. De esta forma, permite a la alta dirección de la empresa tomar decisiones en función de los resultados obtenidos y de su apetito de riesgo.

Cod. Type	Type	Cod. Threat	Threat	Ocurrence Probability	Degradation Percentage
TT3	Nefarious activities or abuse	TT3.4	Abuse of authorizations	100.0	60
TT3	Nefarious activities or abuse	TT3.11	Identity theft (Identity Fraud/ Account)	40.0	20
TT4	Legal	TT4.3	Unauthorized use of IPR protected resources	20.0	40
TT3	Nefarious activities or abuse	TT3.18	Unauthorized activities	40.0	20
TT3	Nefarious activities or abuse	TT3.19	Unauthorized installation of software	60.0	20
TT1	Unintentional damage	TT1.5	Loss of information in the cloud	40.0	100
TT2	Illegal telephone tapping, interception or hijacking	TT2.1	Network Reconnaissance, Network traffic manipulation and Information gathering	20.0	20
TT1	Unintentional damage	TT1.7	Damage caused by a third party	60.0	60
TT2	Illegal telephone tapping, interception or hijacking	TT2.7	Man in the middle/ Session hijacking	40.0	60

Fig. 8. Amenazas con la probabilidad de ocurrencia y el porcentaje de degradación en la herramienta eMARISMA para el caso de estudio.

Cod. Threat	Threat	Type	Asset	Description	P. of Occurrence	DIM3	DIM4	DIM5	DIM1	DIM6	DIM2	Edit
TT3.11	Identity theft (Identity Fraud/ Account)	Data	Databases	Databases	20	-	20	-	-	-	-	
TT3.12	Hoax	Infrastructure	Operating Systems	Operating Systems	20	-	20	-	-	-	-	
TT3.15	Manipulation of hardware and software	Big Data Analytics	Models definitions	Models definitions	20	40	-	-	20	-	20	
TT3.1	Abuse of Information	Infrastructure	Operating Systems	Operating Systems	20	-	-	-	20	20	-	
TT3.2	Social Engineering	Data	Databases	Databases	20	-	100	-	-	-	-	
TT1.12	Destruction of records	Data	Identification record data	Identification record data	20	20	-	-	20	-	-	

Fig. 9. Porcentaje de degradación para las dimensiones afectadas por los activos y amenazas de eMARISMA para nuestro caso de estudio.

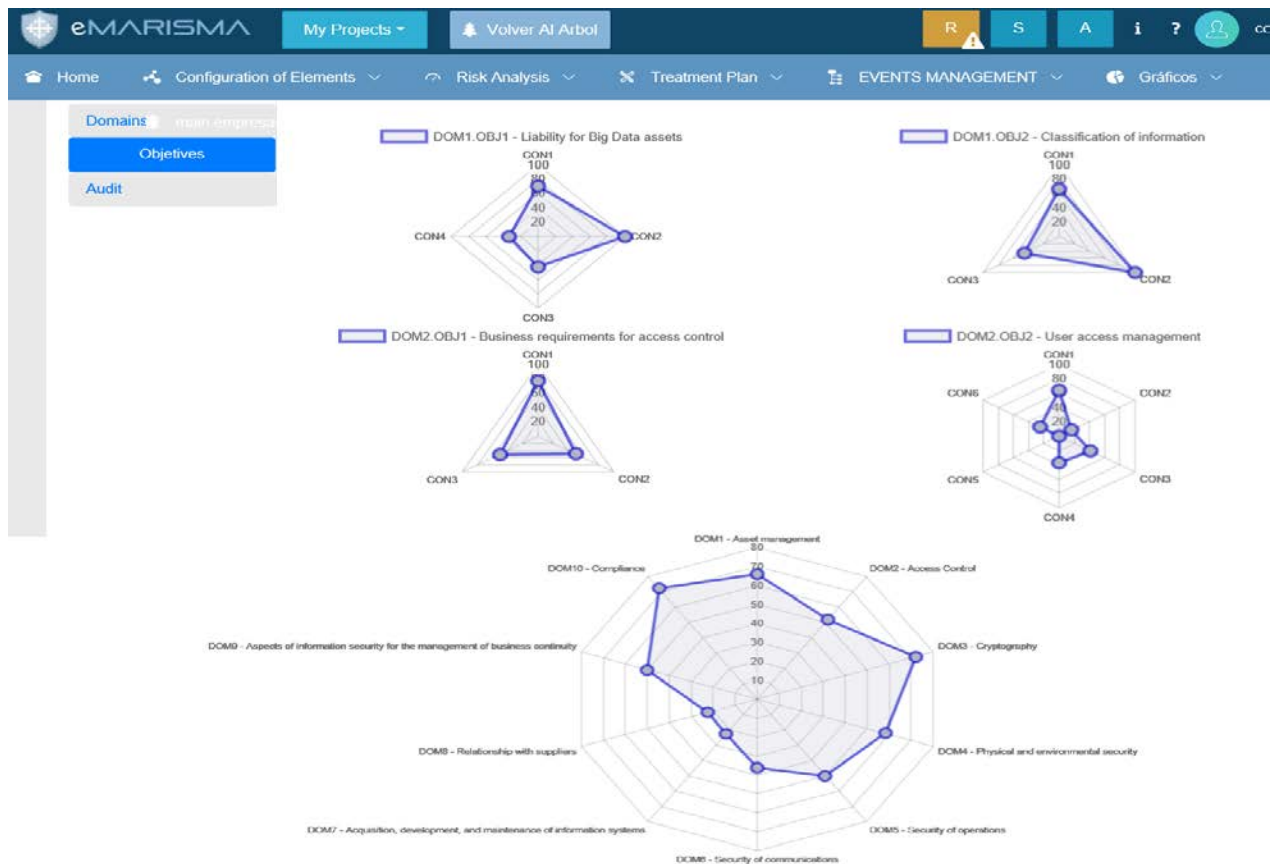


Fig. 10. Diagramas Kiviati generados por la herramienta eMARISMA para los objetivos y auditoría alcanzados en nuestro caso de estudio.

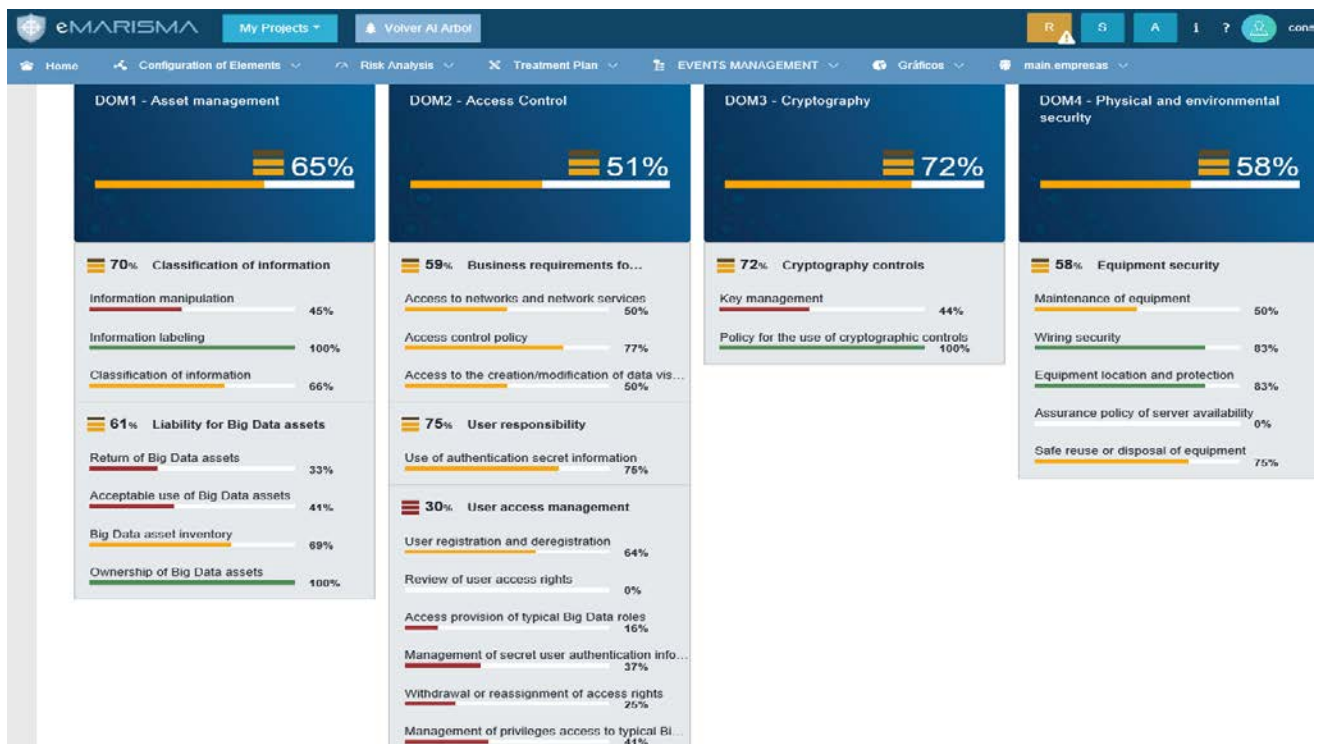


Fig. 11. Cuadro de mando generado por la herramienta eMARISMA para nuestro caso de estudio.

V. CONCLUSIONES Y TRABAJO FUTURO

Un proceso de evaluación y gestión de riesgos puede dividirse en varias etapas, incluyendo el establecimiento del contexto, la identificación de riesgos, el análisis de riesgos en términos de probabilidad e impacto, la evaluación de riesgos y, finalmente, el tratamiento de riesgos.

Este trabajo muestra cómo se utiliza la metodología MARISMA (apoyada por la herramienta eMARISMA), para generar un patrón de gestión y análisis de seguridad enfocado en aspectos de Big Data, que permita una gestión dinámica del riesgo asociado a los elementos de un entorno de Big Data en una empresa.

Esta propuesta se ha aplicado en un caso de estudio, cuya aplicación ha permitido afinarla y validarla con esa experiencia. Estos refinamientos se han centrado principalmente en ajustar los principales conceptos del patrón MARISMA-BiDa, reafirmando los conceptos más relevantes de los ya identificados, y encontrando otros basados en la experiencia. Como trabajo futuro, se contempla la evolución de la herramienta eMARISMA como un sistema de aprendizaje en la nube. Permitirá incorporar los incidentes de seguridad que afecten a uno de los sistemas, en todos aquellos sistemas que estén relacionados o puedan verse afectados.

AGRADECIMIENTOS

Este trabajo ha sido financiado por el proyecto ECLIPSE (Ministerio de Economía, Industria y Competitividad de España y el Fondo Europeo de Desarrollo Regional FEDER, RTI2018-094283-B-C31), y el proyecto GENESIS (Consejería de Educación, Cultura y Deportes de la Dirección General de Universidades, Investigación e Innovación de la Junta de Comunidades de Castilla-La Mancha, España, SBPLY/17/180501/000202). Agradecemos la ayuda de las compañías Sicaman Nuevas Tecnologías SL (www.sicaman-nt.com) y Marisma Shield SL (www.emarisma.com), que han facilitado el uso de la herramienta eMARISMA.

REFERENCIAS

- [1] ENISA, "Good Practices and Recommendations on the," 2015.
- [2] K. Armstrong, "Big data: a revolution that will transform how we live, work, and think," *Information, Commun. Soc.*, vol. 17, no. 10, pp. 1300–1302, Nov. 2014.
- [3] D. Broeders, E. Schrijvers, B. van der Sloot, R. van Brakel, J. de Hoog, and E. Hirsch Ballin, "Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data," *Comput. Law Secur. Rev.*, vol. 33, no. 3, pp. 309–323, Jun. 2017.
- [4] D. Laney, "3D data management: Controlling data volume, velocity and variety," *META Gr. Res. note*, vol. 6, no. 7, p. 1, 2001.
- [5] S. Klous, "Sustainable harvesting of the Big Data potential," *Explor. Boundaries Big Data*, p. 27, 2016.
- [6] B. Hopkins and B. Evelson, "Expand your Digital Horizon with Big Data," *Forrester*, vol. 30, 2011.
- [7] F. J. Alexander, A. Hoisie, and A. Szalay, "Big Data," *Comput. Sci. Eng.*, vol. 13, no. 6, pp. 10–13, Nov. 2011.
- [8] J. Dijcks, "Oracle: Big data for the enterprise," *Oracle White Pap.*, no. June, p. 16, 2012.
- [9] E. Dumbill, "Making sense of big data (editorial)," *Big Data*, vol. 1, no. 1, 2013.
- [10] P. Zikopoulos and C. Eaton, *Understanding big data: Analytics for Enterprise Class Hadoop and Streaming*, vol. 11, no. 1. McGraw-Hill Osborne Media, 2016.
- [11] R. Akerkar et al., "Understanding and mapping big data," no. March. D1, 2015.
- [12] S. V. Bharathi, "Prioritizing and Ranking the Big Data Information Security Risk Spectrum," *Glob. J. Flex. Syst. Manag.*, vol. 18, no. 3, pp. 183–201, Sep. 2017.
- [13] B. Goswami and P. K. Chandra, "Risk Assessment and Analysis for Big Data," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 5, no. 11, 2015.
- [14] A. M. Barrientos and K. A. Areiza, "Integration of a safety management system with an information quality management system," *Universidad EAFIT*, 2005.
- [15] R. Fredriksen, M. Kristiansen, B. A. Gran, K. Stølen, T. A. Opperud, and T. Dimitrakos, "The CORAS Framework for a Model-Based Risk Management Process," *LNCS 2434*, 2002, pp. 94–105.
- [16] G. Disterer, "ISO/IEC 27000, 27001 and 27002 for Information Security Management," *J. Inf. Secur.*, vol. 04, no. 02, pp. 92–100, 2013.
- [17] J. Moreno, L. E. Sánchez, A. S. Olmo, D. G. Rosad, M. A. Serrano, and E. F. Medina, "Marisma-BiDa: Entorno Integrado de Análisis y Gestión de Riesgos en Big Data," in *Actas de las Cuartas Jornadas Nacionales de Investigación en Ciberseguridad*, 2018, pp. 159–165.
- [18] L. Ortiz Restrepo, V. Duque, and F. Javier, "Gestión de riesgos en eTOM. Un análisis comparativo con los estándares de riesgo corporativo," *Rev. Logos, Cienc. Tecnol.*, vol. 9, no. 1, pp. 85–99, 2017.
- [19] L. D. Bodin, L. A. Gordon, and M. P. Loeb, "Information security and risk management," *Commun. ACM*, vol. 51, no. 4, pp. 64–68, Apr. 2008.
- [20] MAGERIT v.3: Metodología de Análisis y Gestión de Riesgos de los Sistemas de Información," 2012.
- [21] G. Wangen, "Information Security Risk Assessment: A Method Comparison," *Computer (Long Beach, Calif.)*, vol. 50, no. 4, pp. 52–61, Apr. 2017.
- [22] C. Fakrane and B. Regragui, "INTERACTIONS AND COMPARISON OF IT RISK ANALYSIS METHODS," in *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech)*, 2018, pp. 1–7.
- [23] F. Jeannot, "Méthodologies d'évaluation et gestion de risques en sécurité," *Montréal, Canada, Mai 2018*, R518, v1.0, 2018.
- [24] A. Benavides, "Modelo de Sistema de Gestión de Seguridad de la Información Basado en la Norma NTC ISO/IEC 27001 para Instituciones Públicas de Educación Básica de la Comuna Universidad de la Ciudad de Pereira," no. 6, pp. 67–72, 2017.
- [25] Werner George Bornman, "Information Security Risk Management: a Holistic Framework," *University of Johannesburg*, 2004.
- [26] A. Refsdal, B. Solhaug, and K. Stølen, "Cyber-risk management," in *Cyber-Risk Management*, Springer, 2015, pp. 33–47.
- [27] R. Zudin, "Analysis of information risk management methods," *Univ. Jyväskylä*, 2014.
- [28] J. P. Carrillo Sánchez, "Guía y análisis de gestión de riesgos en la adquisición e implantación de equipamiento y servicios de tecnologías de información y comunicaciones para proyectos de alcance nacional," *Quito: EPN*, 2012, 2012.
- [29] D. C. Pacheco Pozo, "Propuesta de un plan de contingencia de TI para la empresa LOGICIEL," *Quito*, 2016., 2016.
- [30] L. E. Sánchez, A. S. O. Parra, D. G. Rosado, and M. Piattini, "Managing Security and its Maturity in Small and Medium-sized Enterprises," *J. Univers. Comput. Sci.*, vol. 15, no. 15, pp. 3038–3058, 2009.
- [31] A. Santos Olmo Parra, L. E. Sanchez Crespo, E. Alvarez, M. Huerta, and E. Fernandez Medina Paton, "Methodology for Dynamic Analysis and Risk Management on ISO27001," *IEEE Lat. Am. Trans.*, vol. 14, no. 6, pp. 2897–2911, Jun. 2016.
- [32] J. Pirrone and M. Huerta, "Security Mechanism for Medical Record Exchange Using Hippocratic Protocol," in *IFMBE Proceedings*, vol. 68, no. 1, Springer Verlag, 2019, pp. 401–404.
- [33] T. Vivas, A. Zambrano, and M. Huerta, "Mechanisms of security based on digital certificates applied in a telemedicine network," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 1817–1820.
- [34] A. Santos-Olmo, L. Sánchez, D. Rosado, E. Fernández-Medina, and M. Piattini, "Applying the Action-Research Method to Develop a Methodology to Reduce the Installation and Maintenance Times of Information Security Management Systems," *Futur. Internet*, vol. 8, no. 3, p. 36, Jul. 2016.
- [35] ISO 31000, "ISO 31000:2018 Risk Management," *ISO*, 2018.

- [36] R. Kelemen, M. Biskup, and N. B. Redep, "The conceptual Risk Management Model — A case study of Varazdin County," in 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2016, pp. 1539–1545.
- [37] M. A. Khan, M. F. Uddin, and N. Gupta, "Seven V's of Big Data understanding Big Data to extract value," in Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education, 2014, pp. 1–5.
- [38] Z. Sun, K. Strang, and R. Li, Big Data with Ten Big Characteristics. 2019.
- [39] M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," Mob. Networks Appl., vol. 19, no. 2, pp. 171–209, Apr. 2014.
- [40] NIST, "NIST Big Data Interoperability Framework: Volume 1, Definitions, version 2," NIST Special Publication 1500-1r1, Jun-2018. [Online]. Available: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-4r1.pdf>.
- [41] E. Rekleitis, "Big Data Threat Landscape and Good Practice Guide," Eur. Union Agency Netw. Inf. Secur., no. January, 2016.
- [42] P. Murthy, A. Bharadwaj, P. Subrahmanyam, A. Roy, and S. Rajan, "Big Data Taxonomy," Cloud Security Alliance, no. September. Cloud Security Alliance, September, p. 33, 2014.
- [43] ENISA, "Threat Landscape and Good Practice Guide for Software Defined Networks/5G - SDN Threat Landscape," Jan-2016. .
- [44] NIST, "NIST Big Data Interoperability Framework: volume 3, use cases and general requirements," NIST Special Publication 1500-3r1, Jun-2018. [Online]. Available: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-3r1.pdf>.

David G. Rosado tiene un Máster y es doctor en Informática por la Universidad de Málaga y por la Universidad de Castilla-La Mancha, respectivamente. Es profesor titular en la Escuela Superior de Informática de la Universidad de Castilla-La Mancha en Ciudad Real. Su actividad investigadora se centra en seguridad de sistemas de información, Cloud Computing y Big Data. Sobre estos temas, él ha publicado numerosos artículos en conderencias de ámbito nacional e internacional, también es editor y coeditor de varios libros. Es autor de varias publicaciones en revistas nacionales e internacionales (Information Software Technology, System Architecture, Network and Computer Applications, etc.). Él es miembro del comité de programa de numerosas conferencias y workshops nacionales e internacionales. Es miembro del grupo de investigación GSyA del departamento de sistemas de información y tecnologías de la Universidad de Castilla-La Mancha. Su correo electrónico es david.grosado@uclm.es.

Julio Moreno es Máster y estudiante de doctorado en Informática por la Universidad de Castilla-La Mancha. Su investigación se centra en la seguridad y privacidad de los datos, así como en la creación de un marco de gobierno para Big Data. Es miembro del grupo de investigación GSyA del departamento de sistemas de información y tecnologías de la Universidad de Castilla-La Mancha. Su correo electrónico es julio.moreno@uclm.es.

Luis Enrique Sánchez es Doctor y Máster en Informática y es profesor ayudante doctor de la Universidad de Castilla-La Mancha (Ciudad Real, España), Master en Auditoría de Sistemas de Información por la Universidad Politécnica de Madrid, y Auditor Certificado de Sistemas de Información por ISACA. Es Director de los departamentos de Servicios Profesionales y de I+D de la empresa Sicaman Nuevas

Tecnologías S.L. Sus actividades de investigación son sistemas de seguridad de gestión, métricas de seguridad, minería de datos, limpieza de datos e inteligencia de negocios. Participa en el grupo de investigación GSyA del departamento de sistemas de información y tecnologías de la Universidad de Castilla-La Mancha, en Ciudad Real (España). Su correo electrónico es luisenrique@sanchezcrespo.org.

Antonio Santos-Olmo es Licenciado en Informática y es profesor asociado de la Escuela Superior de Informática de la Universidad de Castilla-La Mancha en Ciudad Real (España), Master en Auditoría de Sistemas de Información por la Universidad Politécnica de Madrid, y Auditor Certificado de Sistemas de Información por ISACA. Es Director de los departamentos de Software Factory de la empresa Sicaman Nuevas Tecnologías S.L. Sus actividades de investigación son sistemas de seguridad de gestión, métricas de seguridad, minería de datos, limpieza de datos e inteligencia de negocio. Participa en el grupo de investigación GSyA del Departamento de de sistemas de información y tecnologías de la Universidad de Castilla-La Mancha, en Ciudad Real (España). Su correo electrónico es asolmo@sicaman-nt.com.

Manuel A. Serrano es Máster y doctor en Informática por la Universidad de Castilla-La Mancha. Es profesor titular en la Escuela Superior de Informática de la Universidad de Castilla-La Mancha en Ciudad Real. Su investigación se centra en la calidad de sofotware de los datos, la medición de software y la calidad y medición de almacenes de datos y Big Data. Su correo electrónico es manuel.serrano@uclm.es.

Eduardo Fernández-Medina es Máster y doctor en Informática por la Universidad de Castilla-La Mancha. Es profesor catedrático en la Escuela Superior de Informática de la Universidad de Castilla-La Mancha en Ciudad Real (España). Su actividad investigadora se centra en el campo de la seguridad de sistemas de información, en particular en seguridad en Big Data, Cloud Computing y sistemas ciberfísicos. En estas temáticas, él es coeditor de varios libros y capítulos de libros, y ha publicado numerosos artículos en conferencias nacionales e internacionales (BPM, UML, ER, ESORICS, TRUSTBUS, etc.). Es autor de más de cincuenta publicaciones en revistas internacionales (Decision Support Systems, Information Systems, ACM Sigmod Record, Information Software Technology, Computer & Security, Computer Standards and Interfaces, etc.). Él lidera el grupo de investigación GSyA del departamento de sistemas de información y tecnologías de la Universidad de Castilla-La Mancha y pertenece a varias asociaciones profesionales y de investigación (ATI, AEC, AENOR, etc.). Su correo electrónico es eduardo.fdezmedina@uclm.es.

ANÁLISIS FORENSE APLICADO A SISTEMAS MULTIMEDIA

Detección de Manipulaciones Copy-Move en Ficheros Multimedia mediante la Transformada Discreta del Coseno

Esteban Alejandro Armas Vega, Ana Lucila Sandoval Orozco, and Luis Javier García Villalba, *Member, IEEE*

Resumen—Las imágenes digitales tienen un papel muy importante en la vida cotidiana. La mayoría de la población tiene una cámara fotográfica de última generación integrada en su dispositivo móvil. El desarrollo tecnológico no sólo facilita la generación de contenido multimedia, sino también la manipulación intencionada de éste, y es aquí donde las técnicas forenses de detección de manipulaciones sobre imágenes cobran gran importancia. En este trabajo se propone una técnica forense basada en el algoritmo de compresión para detectar alteraciones de tipo *copy-move* en una imagen, utilizando para ello la transformada discreta del coseno. Las características obtenidas de estos coeficientes permite obtener vectores de transferencia, los cuales se agrupan y mediante el uso de un umbral de tolerancia permite determinar si existe o no regiones duplicadas dentro de la imagen analizada. Los resultados obtenidos de los experimentos llevados a cabo en este trabajo demuestran la eficacia del método propuesto. Para la evaluación del método propuesto se realizaron experimentos con bases de datos públicas de imágenes falsificadas que son ampliamente utilizadas en la literatura.

Palabras Clave—Análisis Forense, Copia-pegar, Imágenes Digitales, Manipulación, Transformada Discreta del Coseno.

I. INTRODUCCIÓN

El uso de dispositivos móviles ha aumentado considerablemente convirtiéndose en una herramienta que forma parte de la vida cotidiana de la sociedad actual. En 2017, un informe de Cisco Systems [1] indica que el tráfico de datos móviles se ha multiplicado por 18 en los últimos 5 años y se espera que este tráfico continúe aumentando. Estos datos fueron confirmados en 2018 por Ericsson [2] que estima que para el año 2023 el tráfico de datos móviles se multiplicará por 7 y casi tres cuartos del tráfico de datos móviles del mundo se utilizará para transferencia de ficheros multimedia y redes sociales. Como consecuencia, el proceso de compartir datos de forma masiva es fácil y casi inmediato. Las imágenes y vídeos digitales son, gracias a las redes sociales y a las aplicaciones de mensajería instantánea, uno de los recursos que más tráfico de datos genera actualmente.

Por otro lado, la continua mejora de las prestaciones de las cámaras incorporadas en los dispositivos móviles junto a la evolución de las herramientas de edición de imágenes hacen más sencillo manipular una imagen con excelentes resultados. Para enfrentar este tráfico masivo de imágenes manipuladas el

área de análisis forense investiga nuevas técnicas de detección de manipulaciones, para evaluar la integridad de una imagen.

Las imágenes manipuladas llevan existiendo desde hace muchas décadas y están presentes en muchos sectores (política, cine, prensa, rama judicial, etc.).

La manipulación de contenido visual no ha sido algo exclusivo de la era digital actual. A lo largo del tiempo la manipulación siempre ha estado presente. Una de las primeras imágenes manipuladas de la historia [3], es la del fotógrafo Hippolyte Bayard, quien creó una imagen falsa suya suicidándose. Posteriormente, se descubrió que la fotografía fue hecha por el sentimiento de frustración del autor al perder la oportunidad de convertirse en “el inventor” de la fotografía, en lugar de Louis Daguerre que patentó el proceso fotográfico. En el cuadro “El Juicio Final”, el pintor Miguel Ángel cubrió la desnudez de algunas figuras a posteriori por orden del Papa. En fotografía convencional, era posible la manipulación mediante empalme de los negativos de las fotografías, por ejemplo, en la Figura 1 se muestra la manipulación realizada a la famosa foto del dictador soviético Iósif Stalin con su comisario para Asuntos Internos Nikolai Yezhov (Figura 1(a)) para eliminarlo de la foto por orden de Stalin tras ser ejecutado en 1940 (Figura 1(b)).



(a) Imagen Original

(b) Imagen Manipulada

Figura 1: Ejemplo de Manipulación en Fotografía [4]

La facilidad para manipular imágenes y vídeos digitales se ha incrementado en los últimos tiempos y está al alcance del usuario convencional mediante programas como Adobe Photoshop, GIMP, Adobe Premiere, etc. Manipulaciones como los embellecedores de rostros, cambios de expresión facial, mejora de iluminación de la escena, etc., ya las hace de manera automática nuestro dispositivo móvil mediante nuevas herramientas que hacen uso de inteligencia artificial. Por tanto, detectar imágenes digitales manipuladas es de gran importancia en muchas áreas y con diferentes objetivos. Una de las áreas en donde la verificación de la legitimidad de una imagen es fundamental es en lo judicial, donde las imágenes o vídeos pueden suponer evidencia de gran valor para la

E. A. Armas Vega, A. L. Sandoval Orozco and L. J. García Villalba. Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Ingeniería del Software e Inteligencia Artificial, Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, España. e-mail: esarmas@ucm.es, {asandoval, javiergv}@fdi.ucm.es.

resolución de la demanda. Un ejemplo de esto fue el arresto de un conductor [5] que conducía su automóvil a más de 200 Km/h y la evidencia utilizada por la fiscalía fue el vídeo grabado por un peatón, a través del cual se demostró que el imputado circulaba a dicha velocidad.

Sin embargo, para que una imagen pueda ser usada como prueba válida o evidencia en un juicio, se debe asegurar su integridad y demostrar que no ha sido objeto de manipulación. Para llevar a cabo este tipo de autenticación es necesario hacer uso de técnicas robustas de identificación de manipulaciones que puedan garantizar con gran fiabilidad que la imagen es original. Una imagen puede ser manipulada mediante el uso de una variedad de técnicas de manipulación, como *copy-move*, empalme, retoque, filtrado, etc.

En Julio del año 2017 los investigadores de la revista Cognitive-Research [6] utilizaron un dataset de 40 escenas, 30 de las cuales fueron sometidas a cinco tipos diferentes de manipulación, incluyendo manipulaciones físicamente plausibles y no plausibles. Se mostraron a 707 participantes con el fin de evaluar la capacidad de las personas para detectar escenas manipuladas del mundo real. El estudio encontró que sólo el 60 % de las personas fue capaz de detectar las escenas falsas, e incluso entonces, sólo un 45 % de ellos fueron capaces de decir dónde exactamente se encontraba la alteración del contenido. Por todo lo anterior, se deben estudiar y proponer técnicas forenses que permitan hacer frente al gran número de imágenes manipuladas que existen hoy en día.

El resto del trabajo está organizado como sigue: La Sección II detalla las características de las manipulaciones comúnmente utilizadas. En la Sección III-D se describen las principales técnicas de detección de imágenes manipuladas, haciendo énfasis en las técnicas con enfoque pasivo más relevantes de la literatura. Los detalles de la técnica de detección propuesta en este trabajo se presenta en la Sección IV. En la Sección V se analizan los resultados de los experimentos realizados y, finalmente, las conclusiones del trabajo se recogen en la Sección VI.

II. MANIPULACIÓN DE IMÁGENES

Entre los tipos de manipulación de imágenes, destacan los siguientes: retoque, *Copy – Move*, empalme de imágenes y falsificación de huellas digitales [7].

II-A. *Copy – Move*

La manipulación “*Copy-Move*” típicamente se realiza con el objetivo de hacer que un objeto “desaparezca” de la imagen original cubriéndolo con un pequeño fragmento copiado de otra parte de la misma imagen. Este método también se usa para duplicar objetos existentes en la imagen. Como estos bloques copiados provienen de la misma imagen todas sus características serán compatibles con el resto del contenido por lo cual hace muy difícil que el ojo humano lo detecte. Cuando se pega la región copiada se suele acompañar del efecto “*blurring*” en los bordes del área modificada para disminuir las irregularidades entre la región original y la modificada. Las técnicas de detección de este tipo de manipulación se centran en la búsqueda de áreas duplicadas. Sin embargo, si la manipulación se combina con otras técnicas de post-procesamiento, como la aplicación de filtros de color,

el reconocimiento del área modificada es mucho mas difícil para este tipo de técnicas[8][9]. Un ejemplo de este tipo de manipulación se muestra en la Figura 2. En la Figura manipulada 2(b) se han duplicado los dos animales que aparecían en la Figura 2(a).

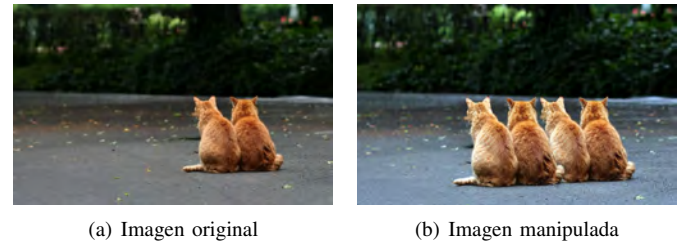


Figura 2: Ejemplo de Manipulación con la Técnica Copia-Pega [10]

II-B. *Empalme*

Esta técnica es similar a la técnica “*Copy-Move*”, con la diferencia de que el fragmento que se copia no pertenece a la misma imagen, es decir, la imagen manipulada es el resultado de la mezcla de dos o más imágenes. El objetivo de esta técnica es insertar elementos que no estaban en la escena que fue capturada originalmente. Por regla general, el bloque de imagen “donante” ha podido ser adquirido por otro dispositivo móvil y por tanto sus características y rastros serán diferentes al resto de la imagen. Es muy usada en fotomontajes donde se combinan dos imágenes dando la sensación de ser una sola. Detectar el área exacta que se ha falsificado en la imagen, mediante la técnica de empalme, es de gran complejidad en comparación con la técnica de manipulación anterior. Esto se debe a que no es posible buscar áreas duplicadas ya que la región manipulada proviene de una imagen diferente [11].

En la Figura 3 se muestra un ejemplo de esta técnica. La Figura 3(a) es la imagen donante, el faro es copiado y pegado en la imagen receptora (Figura 3(b)) el resultado del empalme se muestra en la Figura 3(c).

Las técnicas de detección de empalme se centran en hallar la región de la imagen que contenga estas variaciones de características y rastros con respecto al resto del contenido de la imagen original.

II-C. *Aplicación de Filtros*

Esta técnica de manipulación es de las más utilizadas por su sencillez. Casi todos los programas de edición de imágenes digitales incorporan una selección de filtros ya predefinidos para aplicarlos automáticamente sobre la imagen e incluso los dispositivos móviles con cámaras integradas pueden aplicar este tipo de “mejoras” a la imagen al momento de capturar la escena. La aplicación de filtros tiene como objetivo mejorar el acabado final de la imagen modificando aspectos como tonos, saturaciones, brillos, contrastes, etc. No tienen porqué conllevar un cambio “malicioso” en el contenido de la imagen pero se toma en cuenta porque puede aplicarse cualquiera de estos filtros en combinación con otras técnicas de manipulación de imágenes y es muy probable que esto afecte la precisión de los algoritmos de detección de manipulación en imágenes. En



Figura 3: Ejemplo de Manipulación con la técnica de Empalme [12]

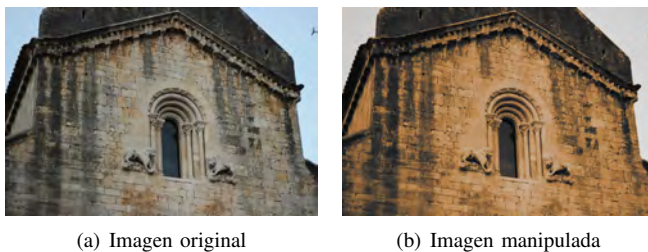


Figura 4: Ejemplo de Manipulación mediante la aplicación de un filtro.



Figura 5: Portada manipulada de la revista Nitro[13]

la Figura 4 se muestra un ejemplo de esta técnica, donde se ha aplicado el filtro predefinido “Sepia” del software de edición GIMP.

II-D. Retoque

Esta manipulación consiste en aplicar pequeñas modificaciones sobre la imagen original sin copiar ningún área del resto de la imagen o tomarla de una diferente. Tiene

como objetivo perfeccionar acabados u ocultar imperfecciones con fines estéticos manteniendo siempre unas características similares a las de la imagen original. Para ello se copian y pegan regiones de la imagen de la misma área. Los retoques que se realizan suelen estar enfocados a perfeccionar la escena [14].

El acabado de las imágenes varía dependiendo del contenido y de los fines con los que se realiza la alteración. Las herramientas más utilizadas en este tipo de manipulaciones suelen ser el saneado, perfilado, emborronado, difuminado y realce. Esta técnica de manipulación es muy común en los sectores de la publicidad, cine y comunicación [15]. La Figura 5 muestra un ejemplo de retoque fotográfico en el que la apariencia de la modelo se modificó digitalmente. La Figura 5(a) muestra la imagen original sin retoque y la Figura 5(b) muestra el resultado de retocar la imagen para la portada de la revista *Nitro*.

Las portadas y los anuncios de las revistas de moda generalmente utilizan algún tipo de retoque para ocultar las imperfecciones y así aumentar los niveles de belleza en las fotografías.

II-E. Manipulación de la Huella Digital

La manipulación de la huella digital de una imagen no está centrado en la parte visual de la imagen si no en la información que ésta contiene [16].

La huella digital es un rastro que dejan todas las cámaras de los dispositivos móviles sobre la imagen que toman durante el proceso de captura. Cuando se genera una imagen digital se introduce este rastro, también llamado ruido.

Extraer el ruido de una imagen proporciona una información valiosa acerca de la fuente (modelo y marca del dispositivo) que generó dicha imagen ya que el tipo de ruido que contiene es intrínseco y único al modelo de la cámara que lo generó. El objetivo de manipular la huella digital de una imagen es el de poder modificar su origen. Si se sustituye la huella digital de la imagen por otra, es posible incriminar a otro dispositivo móvil en la escena en cuestión. También es posible eliminar la huella y así anonimizarla. Este tipo de técnica se subdivide en: Anonimización de la imagen, que consiste en eliminar la información del origen de la imagen y, la falsificación de la imagen, que elimina la huella de la cámara que generó la imagen y coloca una huella de otra cámara de un dispositivo diferente. Estas técnicas no implican la alteración de la imagen en sí, sino la modificación de la información asociada (huella digital) que proviene del sensor que capturó la imagen.

Existen varias fuentes de imperfecciones y ruido introducidas durante el proceso de adquisición de imágenes. Esas imperfecciones aparecen principalmente por dos razones; primero hay componentes aleatorios como el ruido de lectura o el ruido de disparo y segundo debido al ruido del patrón, que es un componente determinista del sensor y permanece aproximadamente igual si varias fotos de la misma escena están tomados. Este patrón es útil para detectar la fuente de origen de una imagen, ya que cada dispositivo tendrá un patrón de ruido específico [16] [17] .

III. TÉCNICAS DE DETECCIÓN DE MANIPULACIONES

Existen dos enfoques forenses de detección de imágenes manipuladas: Intrusivo o activo y no intrusivo o pasivo [18].

- **Enfoque Activo:** Analiza las marcas de agua o señales que deja un dispositivo al momento de generar una imagen digital. El mayor inconveniente de este tipo de enfoque es que muchas cámaras no tienen la capacidad de incorporar este tipo de marcas o firmas, por lo que su alcance es limitado.
- **Enfoque Pasivo:** Analiza el contenido y las características de la imagen digital. A su vez este enfoque puede clasificarse en: métodos basados en aprendizaje y métodos basados en bloques.

El enfoque pasivo tiene un alcance más amplio que el enfoque activo ya que no necesita información previa sobre las imágenes. A continuación se presentan las propuestas de enfoque pasivo más relevantes. La Figura 6 Muestra una clasificación de las técnicas de detección de manipulaciones en imágenes digitales

III-A. Detección de Copy-Move

La detección de falsificaciones “copy-move” son las técnicas más utilizadas en el campo forense debido a su simplicidad y eficacia. La principal evidencia que se explota para detectar este tipo de manipulación es la existencia de dos áreas iguales basándose en las propiedades de los bloques en los que se divide la imagen.

La primera aproximación que se realizó para identificar áreas copiadas fue realizada en el año 2003, en [19] los autores propusieron un método que hacia uso Transformada Discreta del Coseno (DCT) para localizar coincidencias entre bloques de una manera más eficiente que la de realizar una búsqueda por fuerza bruta.

En [20] se propone un método que utiliza el Análisis de Componentes Principales (PCA) para representar una imagen como una representación de bloques superpuestos. Obtuvieron unos resultados más eficientes que los obtenidos en la técnica anterior debido a que consiguieron reducir el coste computacional al rebajar a la mitad el número de cálculos requeridos para procesar los bloques con PCA. Aún así, el coste computacional seguiría siendo grande y por ello, en [21] se propone encontrar las coincidencias entre los bloques buscando patrones de intensidad similares.

Los autores de [22] proponen reducir aún más el tamaño de los bloques superpuestos en los que se subdivide la imagen

para mejorar la precisión al comparar similitudes entre ellos. Las posibles áreas duplicadas tendrán unas propiedades de intensidad similares. Este método es más efectivo frente a pérdidas por compresión que los métodos anteriores.

En [23] se propone detectar las regiones duplicadas estudiando todas las invariantes de desenfoque de una imagen. Los resultados de éste método fueron correctos pero con la desventaja de obtener un tiempo de computación demasiado grande (un promedio de 30 minutos para una imagen RGB de tamaño medio).

En [24] los autores desarrollaron un método que descomponía la imagen en registros de coordenadas polares y, haciendo uso de la Transformada Wavelet, detectar las regiones copiadas. Se reducía así la dimensión de la imagen de entrada debido a la aplicación de Wavelet. Para encontrar los bloques similares se realizaba una búsqueda por fuerza bruta mapeando cada uno de los bloques con las coordenadas polares y la correlación entre ellos como criterio. Otros autores en [25] hicieron algo similar pero se basaron en la Transformada de Fourier (FFT).

En [26] utilizan Transformada Wavelet (WT) para reducir la dimensión de la imagen y Descomposición en Valores Singulares (SVD) para generar el vector de características de cada región con el fin de buscar las similitudes con mayor eficacia. Las regiones duplicadas eran localizadas por clasificación lexicográfica y vecindad detectando todos los bloques, incluso cuando la imagen había sido muy comprimida.

Los métodos descritos hasta ahora no producían resultados óptimos cuando las imágenes sufrían cierta transformación geométrica. En [27] los autores proponen una nueva metodología basada en el algoritmo *Scale-invariant Feature Transform* (SIFT) para estimar los parámetros de la transformación geométrica aplicada sobre la imagen (traslación horizontal o vertical, escalados o rotación del ángulo) con alta fiabilidad, pudiendo así, detectar falsificaciones en imágenes que han sufrido alguna de estas transformaciones. Los autores de [28] mejoran la robustez de SIFT proponiendo un método basado en el algoritmo *Speeded Up Robust Features* (SURF). Este método es además capaz de detectar también áreas copiadas a las que se le ha aplicado modificaciones en brillo o contraste.

III-B. Detección de Empalme

El empalme de imágenes es uno de los esquemas de manipulación más simples y comúnmente utilizados. La detección de este tipo de manipulación es una tarea fundamental durante la verificación de la integridad de imágenes.

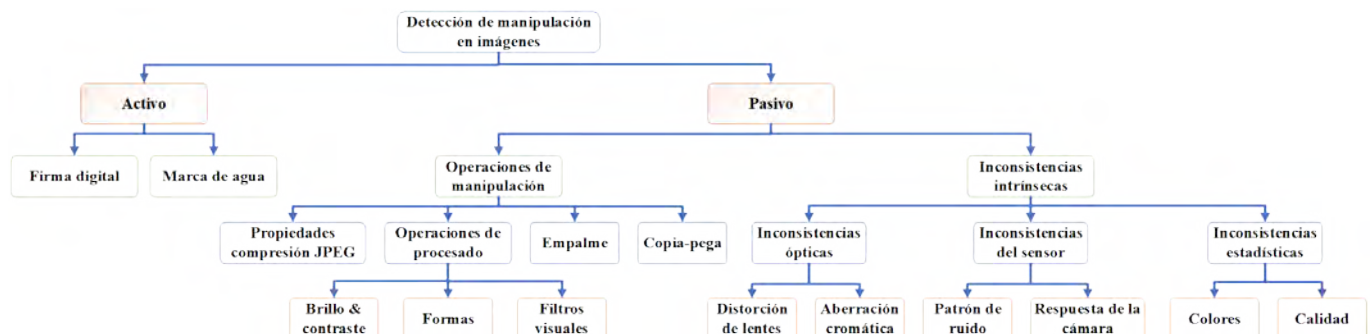


Figura 6: Esquema de detecciones de manipulación en imágenes.

Por lo general, todas las técnicas se basan en las variaciones que se encuentran en el patrón de características del área pegada respecto del contenido de la imagen original.

El primer método propuesto fue presentado en [29]. En este trabajo se propone una técnica basada en el análisis de la señal de la imagen para detectar las correlaciones no naturales que se introducen durante el proceso de falsificación. Obtuvo buenos resultados cuando la detección se llevaba a cabo sobre empalmes realizados por personas y no por máquinas.

En [30] los autores presentaron un modelo de detección de empalme de imágenes basado en el uso de características de magnitud y fase de la propia imagen. Los resultados de la precisión de detección fueron de aproximadamente del 70 %. Posteriormente, los mismos autores propusieron un método para detectar empalmes abruptos utilizando las mismas características.

Los autores de [31], por su parte, se apoyaron en la transformada de Hilbert-Huang para generar estadísticas con el fin de utilizarlas para la clasificación de un modelo de imagen natural. Este modelo estaba basado en los momentos característicos obtenidos con ayuda de la descomposición de Wavelet y así conseguir distinguir las imágenes empalmadas de las imágenes auténticas.

Los autores de [32] propusieron extraer las invariantes de geometría de los píxeles de cada región de una imagen para estimar la Función de Respuesta de la Cámara (CRF) y estudiar las variaciones entre las distintas zonas de la imagen para, así, detectar las áreas que han sufrido un empalme.

En [33] los autores investigaron las características estadísticas de los bloques de una imagen para detectar empalmes. Estas características son extraídas de matrices 2D, las cuales son generadas al aplicar al bloque de imágenes de varios tamaños la DCT. Los experimentos tuvieron una precisión del 91 %.

En [34] se presentó un método de detección basado en características de momento extraídas de la DCT y en métricas de calidad de imagen extraídas de la propia imagen. Descubrieron que ambas características sufrían variaciones cuando una imagen había sufrido un empalme y explotaron dichas variaciones.

Los autores de [35] sugieren un método basado en dividir la imagen por áreas para después extraer las características de densidad de los coeficientes DCT vecinos de cada área. Todas las variaciones en las densidades se clasifican mediante un Máquina de Soporte Vectorial (SVM) para identificar si esas áreas son diferentes.

En [36] se construye un método de detección de empalmes basado en el estudio de las sombras de la imagen. Mediante combinaciones y estimaciones de las zonas de sombra logran encontrar bloques empalmados. Los mismos autores en [37] utilizan la teoría de la homografía plana para localizar la región manipulada y aparte, desarrollaron un método de extracción automático que segmentaba el objeto falso de la imagen manipulada.

En [38] los autores proponen un método basado en la comparación de los espacios cromáticos que forman la imagen. Se utilizan cuatro vectores *Run-length Run-number* (RLRN) con diferentes direcciones extraídos de los canales de crominancia que tienen correlación con características utilizadas en la detección de empalmes.

Los autores de [39] investigaron una técnica basada en el estudio de la iluminación de los objetos de una imagen. Se basan en la consistencia que debe existir en las sombras en función del grado de iluminación y en las características de color del valor de dicha sombra.

En [40] se evaluó la técnica Error Level Analysis (ELA) en imágenes manipuladas con distintos métodos, demostrando una que sólo era efectiva para detectar empalmes.

En [41] se propuso un algoritmo basado en el uso de ELA que demostró detectar con éxito la imagen modificada y el punto exacto de la modificación mediante el uso de histogramas.

En [42] utilizan la Transformada de Wavelet para filtrar los resultados de aplicar ELA con el objetivo de resaltar las alteraciones sobre la imagen.

En [43] se propone un método que combina el descriptor de textura *Local Binary Patterns* (LBP) junto a DCT para detectar cambios producidos por las manipulaciones de empalme y también de copia-pegar. En los experimentos se utiliza SVM obteniendo una tasa de acierto entre el 97.50 % y el 97.77 % sobre el conjunto de datos “CASIA TIDE v2.0”.

El método de detección de empalme propuesto en [44] modela los cambios de manipulación utilizando características estadísticas extraídas de matrices 2D generadas al aplicar la transformada discreta de coseno de bloques de varios tamaños (*Multi-size Block Discrete Cosine Transform* (MBDCT)). En los experimentos se obtuvo un 91.40 % de acierto sobre el conjunto de datos “Columbia” usando SVM.

En [45] exploraron el efecto de diferentes modelos de color en la detección de falsificación de empalme. En este trabajo, se hace una comparación de los modelos cromáticos frente a los modelos *Red-Green-Blue* (RGB) y de luminancia utilizados comúnmente. Se emplean cuatro vectores RLRN con diferentes direcciones extraídas de canales de crominancia correlacionados como características para la detección de empalme en imágenes. Finalmente, se usa SVM como algoritmo clasificador. El conjunto de datos utilizado en los experimentos son “CASIA TIDE v1.0” y “Columbia” con una precisión de 94.7 %.

III-C. Detección de Manipulación de Huella Digital

La huella digital de una imagen identifica su origen y garantiza su integridad. Las técnicas para detectarla se encargarán pues, de estudiar si dicha huella ha sufrido modificaciones, pues en tal caso, es una evidencia de que la imagen en cuestión ha sido manipulada. Las técnicas se basan principalmente en el estudio de los patrones del ruido del sensor que introduce cada cámara en las imágenes que genera durante el proceso de captura de una fotografía.

En [46] se propone un método basado en la extracción de características del ruido de foto-respuesta no uniforme (Patrón de Ruido de Respuesta no Uniforme (PRNU)), junto con un SVM para su clasificación. Este trabajo se utilizó únicamente en dispositivos móviles y se consiguió mostrar que este método consigue buenos resultados cuando se tiene que clasificar una gran cantidad de cámaras fuente.

En [47] se propone combinar dos métodos de detección: el estudio de las imperfecciones del sensor y las *Transformada Discreta de Wavelet* (DWT). Los resultados confirman que

estas dos técnicas juntas ayudan a rastrear con precisión el dispositivo fuente que tomó la imagen, además del modelo y marca de dicho dispositivo.

En [48] se estudian investigaciones recientes en el campo y proponen la mezcla de dos técnicas (imperfecciones del sensor y transformadas wavelet) para obtener una mejor identificación de fuentes de imágenes generadas con dispositivos móviles. Los resultados muestran que las imperfecciones del sensor y las transformadas wavelet pueden servir conjuntamente como buenas características forenses para ayudar a rastrear la cámara fuente de las imágenes producidas por teléfonos móviles. Además, este modelo también permite determinar con gran precisión la marca y el modelo del dispositivo.

Como se puede observar la mayoría de técnicas emplean los patrones de ruido para identificar y extraer el ruido del sensor. Para poder comprobar si se ha llevado a cabo una modificación o eliminación bastaría con comparar la huella digital de la imagen original y de la imagen manipulada. Sin embargo, nuestra propuesta combina el análisis de los patrones de textura locales junto con las características obtenidas de aplicar la transformada discreta de Wavelets junto con la del Coseno a la imagen.

III-D. Detección Manipulación Retoque

En [15] se propuso un algoritmo eficiente diseñado específicamente para predecir la presencia de retoques en imágenes de portadas de revistas. El conjunto de datos de 468 fotos (originales y retocadas) se valoraron entre 1 (muy similar) y 5 (muy diferente) dependiendo de la cantidad de alteración fotográfica. Se calcularon las modificaciones geométricas y fotométricas de cada foto original y retocada y, posteriormente, se extrajeron ocho estadísticas de resumen que incorporan el grado de retoque fotográfico para calcular la correlación con la valoración de cada foto. Se utilizó el algoritmo de Máquina de Soporte Vectorial SVM para determinar el grado de modificación de la imagen. La precisión máxima obtenida con los experimentos fue de 98,75 %.

El algoritmo propuesto en [49], utiliza una red neuronal para extraer características y SVM para clasificar las imágenes en una clase sin retoques o retocada. En los experimentos se utilizó el conjunto de datos “ND-IITD retouched faces” de 325 de caras retocadas y se obtuvo un 87 % de acierto.

En [50] se propone la extracción de las características de color, forma y textura de tres regiones faciales predefinidas. Se utilizaron los conjuntos de datos *YouTube Makeup* (YMU) y *Makeup in the Wild* (MIW) [51] para entrenar y predecir, respectivamente, un sistema SVM con núcleo *Radial Basis Function* (RBF) para clasificarlas. La precisión que se obtuvo fue de un 93 %. Posteriormente, en [52] se propuso un algoritmo más preciso para la detección de maquillaje en los mismos conjuntos de datos utilizando características de textura y forma. La técnica propuesta extrae un vector de características que captura las características de forma y textura de la cara usada como entrada del algoritmo. Se consiguió aumentar la precisión a un 98.5 % usando un clasificador SVM.

IV. MÉTODO PROPUESTO

En esta sección se propone un esquema mejorado de detección de falsificación de copia-pegar basado en el esquema

presentado por primera vez por Fridrich [19]. A continuación se especifican los parámetros de entrada y los resultados que genera el algoritmo tras su ejecución.

- Entrada: Imagen a analizar.
- Salida: Imagen con la región duplicada marcada de un color determinado, de esta forma se puede visualizar claramente el área sobre el que se ha realizado la copia y la región exacta donde se ha pegado.

El primer paso es convertir la imagen que se desea analizar a escala de grises. Para ello se extraen los componentes del canal de luminancia y se representa la imagen con ellos.

A continuación, se establece un tamaño de bloque $B = 8$ para dividir la imagen desde la esquina superior izquierda a la esquina inferior derecha.

Los bloques se superponen con un desplazamiento de un píxel hasta obtener $(M-B+1)(N-B+1)$ bloques superpuestos, siendo M y N las dimensiones de la imagen. El tamaño de bloque B se ha establecido en 8 para conseguir resultados más precisos con un nivel óptimo de ruido.

Seguidamente se extraen las características DCT de cada uno de los bloques. DCT puede eliminar la redundancia entre píxeles adyacentes de manera rápida y efectiva, y tiene propiedades de compactación de energía [53], por ello es razonable adoptar los coeficientes DCT como características de los bloques de imagen. Una propiedad de DCT es que la energía solo se enfoca en los coeficientes de baja frecuencia, es decir, no todos los elementos son igual de importantes, por ello se descartan los coeficientes de alta frecuencia por que solo introducen ruido y pueden dar lugar a errores en procesos posteriores.

Para llevar a cabo este proceso se establece un valor de truncamiento y se realiza a su vez un escaneado en zigzag. El valor de truncamiento k se calcula mediante la ecuación 1 y corresponde a la longitud del vector de características de un bloque. Para establecerlo se fija un factor de truncamiento $f_t (0 < f_t < 1)$.

$$k = \lceil f_t \cdot B^2 \rceil \quad (1)$$

A su vez se realiza un escaneado en *zigzag* sobre el bloque de coeficientes DCT como se muestra en la Figura 7. Este tipo de escaneo permite realizar un recorrido ascendente por los coeficientes de menor a mayor frecuencia y gracias al valor k antes mencionado accede solo a los coeficientes más importantes. La zona verde de la Figura 7 representa el área de coeficientes DCT más significativos.

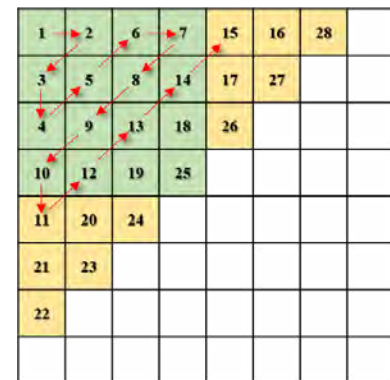


Figura 7: Escaneo en zig-zag de coeficientes DCT

Por último, para reducir las dimensiones y mejorar la eficiencia del proceso de adaptación, los coeficientes DCT se cuantifican mediante un factor de cuantificación f_q y se redondean al entero más cercano usando la ecuación 2.

$$\vec{a}_i = \left(\left\lfloor \frac{a_{i1}}{f_q} \right\rfloor, \left\lfloor \frac{a_{i2}}{f_q} \right\rfloor, \dots, \left\lfloor \frac{a_{ik}}{f_q} \right\rfloor \right) \quad (2)$$

Este proceso proporciona una secuencia con los coeficientes de baja frecuencia agrupados y truncados para cada bloque de la imagen.

$$\#Bloques = (M - B + 1)(N - B + 1)(f_t \cdot B^2) \quad (3)$$

Finalmente, se crea una matriz A de una sola columna para guardar cada secuencia en una fila diferente, a la vez que se añaden las coordenadas x e y de la esquina superior izquierda del bloque al final de la secuencia de coeficientes.

El siguiente paso es ordenar lexicográficamente la matriz A con todos los vectores de características, de esta manera las filas de características similares quedarán juntas y así se podrá determinar qué bloques de la imagen están relacionados. Por lo tanto, se requieren algunos métodos para juzgar si los vectores de características correspondientes de los bloques de imagen son los mismos.

Si los componentes correspondientes de los dos vectores de bloques de imagen son casi iguales, los dos bloques de imagen pueden considerarse estrechamente relacionados. Posteriormente, estos bloques se estudiarán para determinar si uno de ellos o ambos son objetos de manipulación. Para juzgar la similitud entre dos bloques se realizan las siguientes comprobaciones para cada fila de la matriz A :

1. Cada vector de fila $\vec{a}_i = (a_i^1, a_i^2, \dots, a_i^k)$ debe ser comparado con sus vectores de fila adyacentes $\vec{a}_j = (a_j^1, a_j^2, \dots, a_j^k)$.
2. Se define un parámetro N_a que corresponde al número de máximo de filas que van a ser comparadas con a_i , por lo que debe satisfacer $(j-i < N_a)$.
3. Se definen los umbrales S_t y T_t que serán usados más adelante y se inicializa la variable r_{max} a un valor suficientemente pequeño y la variable r_{min} a un valor suficientemente grande.
4. Se crea un contador c inicializado a 0.
5. Para cada $\vec{a}_i$ y $\vec{a}_j$ dentro del intervalo $(1 \leq l \leq k)$ se comprueba que:

- 4.1. Si $a_j^l = 0$ se comprueba si se cumple $|a_i^l - a_j^l| < S_t$, en caso afirmativo se incrementa el valor de c en 1.
- 4.2. Si $a_j^l \neq 0$ se calcula $r_i = a_i^l / a_j^l$ y se cambian los valores r_{min} y r_{max} según corresponda:
 - Si $r_{max} < r_l$ entonces $r_{max} = r_l$
 - Si $r_{min} > r_l$ entonces $r_{min} = r_l$.
- 4.3. Si $r_{max} - r_{min} > T_t$ entonces c se incrementa en 1.
6. Finalmente si $c < C_t$ entonces a_i y a_j son similares.

Después de comprobar si el vector de fila a_i (la coordenada superior izquierda del bloque de imagen es (x_1, y_1)) y el vector de fila a_j (la coordenada superior izquierda del bloque de imagen es (x_2, y_2)) son similares, se calculan los vectores de transferencia entre los dos vectores.

$$\vec{s} = (s_1, s_2) = (x_1 - x_2, y_1 - y_2) \quad (4)$$

A continuación, se comprueba si la distancia excede de un parámetro T_d , con la ecuación 5. Si se cumple, la frecuencia existente del vector de transferencia se incrementa en 1, en caso contrario, no se modifica dicha frecuencia.

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} > T_d \quad (5)$$

Una vez obtenidas las frecuencias para los vectores de transferencia se procede a buscar el vector principal de transferencia cuyas frecuencias exceden un umbral T_f . Los bloques de la imagen correspondientes al vector principal se pueden considerar como regiones copiadas y pegadas. Estas regiones se marcan respectivamente en un color que las haga distinguirse del resto de la imagen. La Figura 8 presenta el diagrama de los procesos más característicos del algoritmo.

V. EXPERIMENTOS Y RESULTADOS

En los experimentos realizados se ha utilizado *Python* como lenguaje de programación, debido a su gran flexibilidad para poder realizar el análisis de datos y su alta velocidad en gestionar la entrada y salida. Para la evaluación del algoritmo se ha hecho uso de varios datasets públicos ([12], [54]) para realizar experimentos con varios formatos y tamaños. La Tabla I muestra las características de los datasets utilizados en los experimentos.

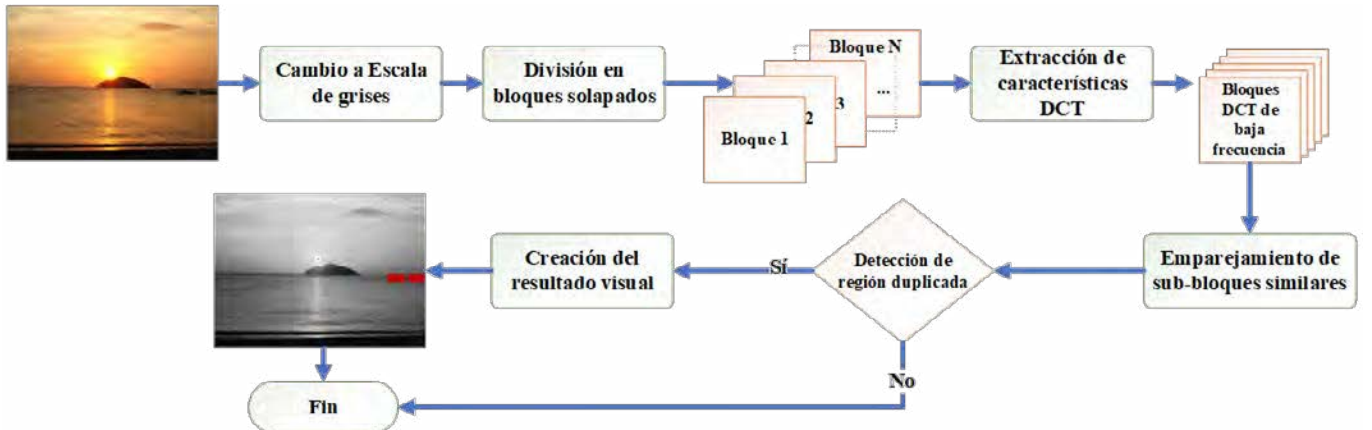


Figura 8: Diagrama de flujo del algoritmo de identificación de Copia-Pega propuesto

Tabla I: Características de los datasets utilizados

Datasets	Formato	Resolución	Número de Imágenes		
			Originales	Manipuladas	Total
CASIA v1.0 [12]	JPEG	384x256	800	921	1721
CASIA v2.0 [12]	JPEG, BMP, JTIFF	240x160 900x600	7491	5123	12614
IFS-TC [54]	PNG	1024x768 3648x2736	424	451	875

Las características del equipo en el cual se han realizado los experimentos se presentan en la Tabla II. Es un factor importante a tener en cuenta ya que los tiempos de ejecución de las diferentes pruebas varían según los recursos computacionales disponibles.

Tabla II: Características del equipo de experimentación

Recursos	Características
Sistema operativo	Ubuntu 17.04
Memoria	12 GB
Procesador	Intel® Core™ i5-6200U CPU @ 2.30GHz x 4
Gráficos	Intel® HD Graphics 520 (Skylake GT2)
Tipo de SO	64 bits
Disco	64 GB

V-A. Experimento 1

El primer conjunto de experimentos se basó en comprobar la efectividad del algoritmo propuesto en la Sección IV.

Este algoritmo hace uso de diferentes parámetros configurables, dependiendo del valor asignado los resultados pueden variar notablemente. En [55] se propone un algoritmo que da excelentes resultados en identificación de manipulaciones copia-pegar. Para realizar sus experimentos hacen comparaciones entre los parámetros usados por otras investigaciones. Los valores que han establecido han servido como referencia para inicializar los parámetros del algoritmo que se ha propuesto en este trabajo. En la siguiente tabla se exponen cada uno de los parámetros utilizados y sus valores correspondientes.

Tabla III: Parámetros configurables del algoritmo copia-pegar

Parámetro	Nombre	Valor asignado
ft	Factor de truncamiento	0,25
fq	Factor de cuantificación	4
Na	Filas vecinas comparables	3
St	Umbral S	4
Tt	Umbral T	0,06
Ct	Umbral de similitud	3
Tf	Umbral de frecuencia	50
Td	Distancia de los vectores	20

El parámetro que mejora los resultados ha sido el *umbral de frecuencia* o *Tf*. Este parámetro establece el valor con el cual un bloque de la imagen puede considerarse una manipulación válida. Si un bloque aparece varias veces en la imagen como duplicado y dicha frecuencia de aparición supera a la establecida por el umbral *Tf* se considerará que forma parte de la manipulación. Estudiar la frecuencia de aparición de los bloques es posible gracias a la superposición con la que se extraen de la imagen. Cuando este parámetro es



(a) Imágenes originales



(b) Imágenes manipuladas



(c) Resultados con Tf=50



(d) Resultados con Tf=100



(e) Resultados con Tf=150

Figura 9: Identificación de la manipulación copia-pegar

alto los resultados finales son más refinados, eliminando las áreas identificadas como manipuladas pero que en realidad son falsos positivos. En el experimento se ha ajustado el parámetro *Tf* en tres valores: 50, 100 y 150. En la Figura 9 se muestran 3 imágenes manipuladas de ejemplo. En la Figura 9(a) se muestran las imágenes manipuladas, en la Figura 9(b) su respectiva imagen sin manipulación y en la Figura 9(c) la zona manipulada.

En la Figura 10 se muestran los resultados de la detección con diferentes valores del parámetro *Tf*. Como se puede observar en la figura a mayor valor del parámetro *Tf* los resultados presentan menos ruido, es decir, se eliminan las zonas negras que no forman parte de la manipulación. En la primera imagen la manipulación se identifica con el *Tf*=50, con un valor más alto el algoritmo no encuentra ningún bloque duplicado que cumpla la frecuencia de aparición establecida por *Tf*. En cambio puede apreciarse en las otras dos imágenes que a mayor valor del parámetro *Tf* se elimina el ruido producido por los falsos positivos. Esto se debe a que son manipulaciones de gran tamaño en proporción a la imagen por lo que la frecuencia de aparición de los bloques manipulados será muy superior al existir la superposición. Sin embargo, el

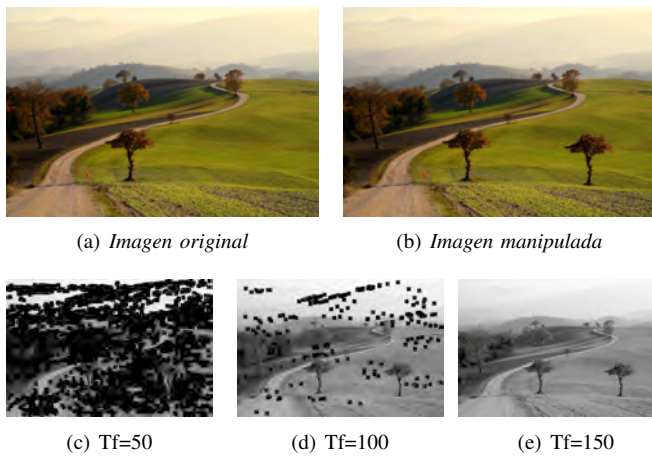


Figura 10: Área duplicada con detalles de la imagen real

algoritmo falla con un tipo concreto de manipulaciones. Estas manipulaciones consisten en duplicar determinadas áreas que presentan zonas de la imagen real. Esto se puede observar en el ejemplo de la Figura 11.

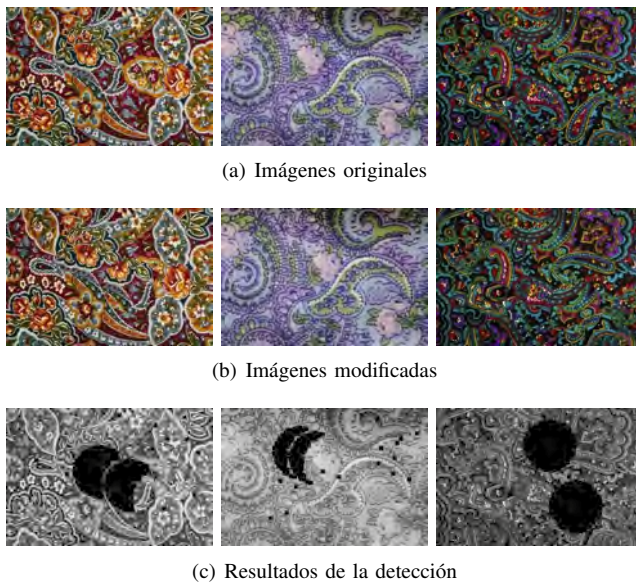


Figura 11: Detección de copia-pegar en regiones de texturas similares

En esta imagen se ha duplicado el árbol situado en la parte central. Este árbol presenta huecos entre las ramas que han sido editados en la duplicación para que se integre perfectamente con el fondo, es por ello que el algoritmo trata ambos árboles como objetos diferentes y no es capaz de dar un resultado correcto.

V-B. Experimento 2

En el segundo experimento se comprobó la precisión del algoritmo de identificación de la región de copia-pegar en imágenes de texturas con patrones similares. En este tipo de imágenes la manipulación pasa inadvertida debido a su excelente integración con el fondo original. Esto se debe a que se usa un mismo patrón de colores sin áreas que resalten por encima de otras. En este experimento se realizaron dos

pruebas con este tipo de imágenes, se ajustó el parámetro Tf al valor 150 para disminuir el ruido de puntos negros en los resultados.

En la primera prueba se usaron imágenes con muchos detalles de múltiples colores pero a su vez siguen un mismo patrón, esto hace que el área duplicada sea difícil de detectar. En la Figura 12 se muestran tres ejemplos de identificación en este tipo de imágenes. Como se observa, el algoritmo consigue una precisión destacable.

Para la segunda prueba se usaron imágenes donde la región duplicada se encontraba en un área del mismo color que otras regiones de la imagen. En este tipo de imágenes también es difícil detectar la región duplicada ya que puede confundirse con otra región original que tenga el mismo color. En la Figura 13 se muestran tres ejemplos donde puede apreciar que el algoritmo presenta un buen funcionamiento ante este tipo de manipulaciones.

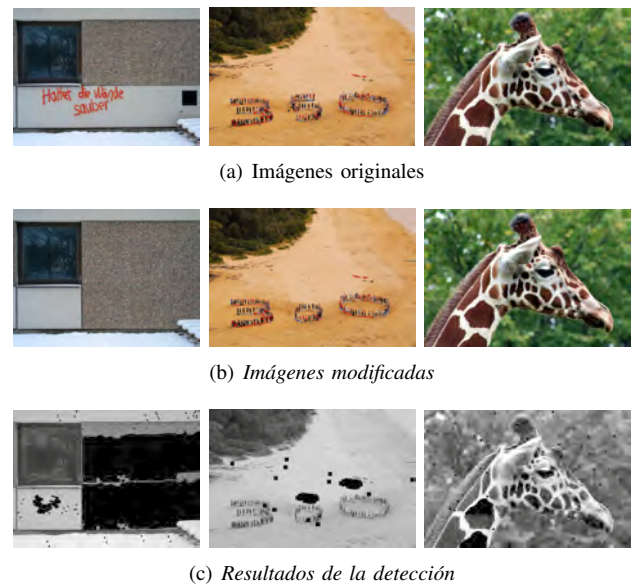


Figura 12: Detección de copia-pegar en imágenes con áreas del mismo color

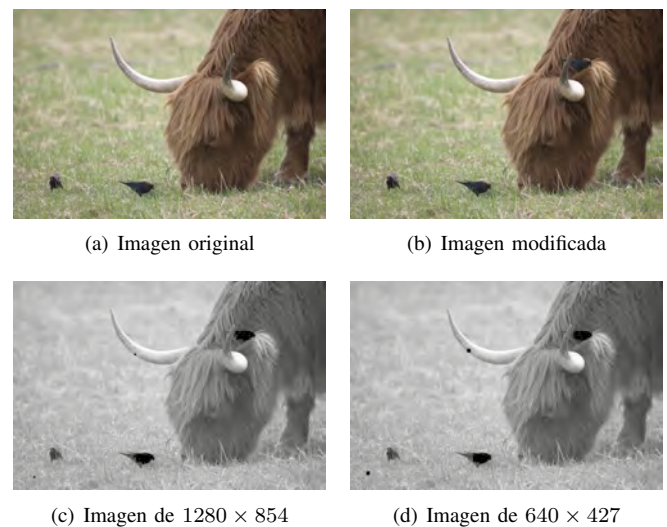


Figura 13: Detección de copia-pegar en imágenes escaladas

V-C. Experimento 3

En este experimento se analizó la eficiencia del algoritmo en imágenes de gran tamaño y resolución. Se observó que al escalar una imagen a un tamaño más pequeño la precisión del algoritmo sigue manteniéndose alta sin sufrir cambios significativos. Esta observación permite realizar un escalado de las imágenes grandes antes de que el algoritmo las procese, esto aumenta la eficiencia sin perder calidad en los resultados.

En la Figura 13 se muestra un ejemplo de una imagen manipulada por la técnica copia-pegar, en la cual se ha copiado el pájaro situado encima del césped y se ha pegado sobre la cabeza de la vaca. El tamaño original de la imagen es de 1080x854 píxeles, también se muestra el resultado obtenido al escalar la imagen a un tamaño de 640x427 píxeles. El tiempo de ejecución que ha tardado el algoritmo en procesar la imagen original ha sido de 160 segundos, en cambio en la imagen escalada ha tardado 48 segundos. Como puede observarse se ha detectado la manipulación perfectamente en ambas imágenes, por lo que es posible realizar el escalado sin afectar a la precisión del algoritmo y mejorando considerablemente el tiempo de ejecución. A lo largo de las pruebas realizadas se ha podido comprobar que el algoritmo funciona con cualquier tipo de formato, como *Joint Photographic Experts Group* (JPEG), *Portable Network Graphics* (PNG), Mapa de Bits (BMP), entre otros. También hay que destacar que el tamaño de la imagen no influye en la precisión de los resultados, solo produce variaciones en el tiempo de procesamiento como se muestra en el Experimento 3.

VI. CONCLUSIONES

Las imágenes digitales contienen una gran cantidad de información relevante. Debido a esto, son un elemento muy importante en el ámbito legal y se han convertido en evidencias que aportan gran valor en la resolución de un juicio. Para que estas evidencias lleguen a ser válidas se debe poder garantizar su autenticidad e integridad de forma fiable. Existen numerosas aplicaciones que consiguen editar imágenes con resultados altamente profesionales y detectar si una imagen ha sido modificada mediante alguna técnica de manipulación es una tarea complicada. Para poder garantizar la integridad de una imagen es de mucho interés tener herramientas forenses que puedan detectar estas falsificaciones. En este trabajo se ha realizado un estudio exhaustivo sobre las técnicas existentes de detección de manipulaciones haciendo énfasis en las técnicas de detección de *Copy – Move*. Se han estudiado en profundidad las técnicas que dan los mejores resultados, analizando el proceso que realiza para la detección. Se ha diseñado una técnica para la detección de la región exacta duplicada en técnicas *Copy – Move*. Para evaluar la técnica diseñada en este trabajo, se han realizado pruebas con un conjunto numeroso de imágenes de diferentes texturas, dimensiones y formatos. Los resultados muestran que el algoritmo detecta de forma precisa las áreas duplicadas en imágenes de texturas similares. De igual forma, muestra excelentes resultados cuando la imagen contiene un alto nivel de detalles similares representados por un mismo patrón de color. Este algoritmo presenta un tiempo de ejecución superior a los demás ya que hace uso de numerosos cálculos para determinar la zona duplicada. Pero como se ha observado en los experimentos es posible

escalar una imagen y así mejorar la eficiencia del mismo sin perder calidad en los resultados. Sin embargo, el algoritmo presenta dificultades en imágenes cuya área duplicada ha sido modificada posteriormente con partes de la imagen original.

AGRADECIMIENTOS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700326. This paper has also received funding from THEIA (Techniques for Integrity and authentication of multimedia files of mobile devices) UCM project (FEI-EU-19-04).



REFERENCIAS

- [1] CISCO, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021," <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html>, February 2017.
- [2] ERICSSON, "Ericsson Mobility Report," ERICSSON, Tech. Rep., 06 2018. [Online]. Available: <https://www.ericsson.com/assets/local/mobility-report/documents/2018/ericsson-mobility-report-june-2018.pdf>
- [3] M. Sapir, "The impossible photograph: Hippolyte bayard's self-portrait as a drowned man," *MFS Modern Fiction Studies*, vol. 40, no. 3, pp. 619–629, 1994.
- [4] H. Farid, "Creating and Detecting Doctored and Virtual Images: Implications to the Child Pornography Prevention Act." UnDartmouth College, Technical Report, September 2004.
- [5] E. Mundo, "Detenido por Circular a 200 Kilómetros por Hora tras Subir un Vídeo a Redes Sociales," <http://www.elmundo.es/madrid/2017/08/30/59a68f0a468aeb7a658b4607.html>, August 2017.
- [6] S. J. Nightingale, K. A. Wade, and D. G. Watson, "Can people identify original and manipulated photos of real-world scenes?" *Cognitive research: principles and implications*, vol. 2, no. 1, p. 30, 2017.
- [7] M. A. Qureshi and M. Deriche, "A Bibliography of Pixel-Based Blind Image Forgery Detection Techniques," *Signal Processing: Image Communication*, vol. 39, pp. 46–74, 2015.
- [8] M. Boutell and J. Luo, "Beyond pixels: Exploiting camera metadata for photo classification," *Pattern Recognition*, vol. 38, no. 6, pp. 935–946, 2005.
- [9] H. Huang, W. Guo, and Y. Zhang, "Detection of copy-move forgery in digital images using sift algorithm," in *Proceedings of the IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, vol. 2, Wuhan, China, December 2008, pp. 272–276.
- [10] I. Fourandsix Technologies, "Photo Tampering Throughout History," <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html>, December 2017.
- [11] X. Zhao, S. Wang, S. Li, J. Li, and Q. Yuan, "Image splicing detection based on noncausal markov model," in *Proceedings of the IEEE International Conference on Image Processing*, Melbourne, VIC, Australia, September 2013, pp. 4462–4466.
- [12] J. Dong and W. Wang, "CASIA TIDE v1.0 - v2.0," <http://forensics.idealtest.org/>.
- [13] V. Sun, "Photos: 20 More Stars and Celebrities Before and After Photoshop," <http://www.vancouver.sun.com/life/fashion-beauty/Photos+more+stars+celebrities+before+after+Photoshop/7841314/story.html>, July 2014.
- [14] I. T. Young, J. J. Gerbrands, and L. J. Van Vliet, *Fundamentals of image processing*. Delft University of Technology Delft, 1998.
- [15] E. Kee and H. Farid, "A Perceptual Metric for Photo Retouching," *National Academy of Sciences*, vol. 108, no. 50, pp. 19907–19912, November 2011.
- [16] L. J. García Villalba, A. L. Sandoval Orozco, J. Rosales Corripio, and J. Hernández Castro, "A PRNU-based Counter-forensic Method to Manipulate Smartphone Image Source Identification Techniques," *Future Generation Computer Systems*, vol. 76, pp. 418–427, November 2017.

- [17] N. Khanna, A. K. Mikkilineni, G. Chiu, J. P. Allebach, and E. Delp, "Forensic Classification of Imaging Sensor Types," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 6505, no. 65050U, February 2007.
- [18] B. Mahdian and S. Saic, "A Bibliography on Blind Methods for Identifying Image Forgery," *Signal Processing: Image Communication*, vol. 25, no. 6, pp. 389–399, July 2010.
- [19] J. Fridrich, D. Soukal, and J. Lukas, "Detection of Copy Move Forgery in Digital Images," in *Proceedings of the Digital Forensic Research Workshop*, Binghamton, New York, August 2003, pp. 5–8.
- [20] A. C. Popescu and H. Farid, "Exposing Digital Forgeries by Detecting Duplicated Image Regions," *Department of Computer Science*, vol. 646, January 2004.
- [21] A. Langille and M. Gong, "An efficient match-based duplication detection algorithm," in *Computer and Robot Vision, 2006. The 3rd Canadian Conference on*. IEEE, 2006, pp. 64–64.
- [22] W. Luo, J. Huang, and G. Qiu, "Robust detection of region-duplication forgery in digital image," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4. IEEE, 2006, pp. 746–749.
- [23] A. D. Warbhe, R. V. Dharaskar, and V. M. Thakare, "International journal of engineering sciences & research technology block based image forgery detection techniques."
- [24] A. Myna, M. Venkateshmurthy, and C. Patil, "Detection of region duplication forgery in digital images using wavelets and log-polar mapping," in *Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on*, vol. 3. IEEE, 2007, pp. 371–377.
- [25] Q. Wu, S. Wang, and X. Zhang, "Log-polar based scheme for revealing duplicated regions in digital images," *IEEE Signal Processing Letters*, vol. 18, no. 10, pp. 559–562, 2011.
- [26] G. Li, Q. Wu, D. Tu, and S. Sun, "A sorted neighborhood approach for detecting duplicated regions in image forgeries based on dwt and svd," in *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 2007, pp. 1750–1753.
- [27] H. Huang, W. Guo, and Y. Zhang, "Detection of Copy-Move Forgery in Digital Images Using SIFT Algorithm," in *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, vol. 2, December 2008, pp. 272–276.
- [28] X. Bo, W. Junwen, L. Guangjie, and D. Yuewei, "Image Copy-Move Forgery Detection Based on SURF," in *In proceedings of the International Conference on Multimedia Information Networking and Security*, Nanjing, China, November 2010, pp. 889–892.
- [29] H. Farid, "Detecting digital forgeries using bispectral analysis," 1999.
- [30] T.-T. Ng and S.-F. Chang, "A model for image splicing," in *Proceedings of the International Conference on Image Processing, 2004*, vol. 2. IEEE, 2004, pp. 1169–1172.
- [31] D. Fu, Y. Q. Shi, and W. Su, "Detection of image splicing based on hilbert-huang transform and moments of characteristic functions with wavelet decomposition," in *International workshop on digital watermarking*. Springer, 2006, pp. 177–187.
- [32] Y.-F. Hsu and S.-F. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency," in *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE, 2006, pp. 549–552.
- [33] Y. Q. Shi, C. Chen, and W. Chen, "A natural image model approach to splicing detection," in *Proceedings of the 9th workshop on Multimedia & security*. ACM, 2007, pp. 51–62.
- [34] Z. Zhang, J. Kang, and Y. Ren, "An effective algorithm of image splicing detection," in *Computer Science and Software Engineering, 2008 International Conference on*, vol. 1. IEEE, 2008, pp. 1035–1039.
- [35] Q. Liu and A. H. Sung, "A new approach for jpeg resize and image splicing detection," in *Proceedings of the First ACM workshop on Multimedia in forensics*. ACM, 2009, pp. 43–48.
- [36] W. Zhang, X. Cao, J. Zhang, J. Zhu, and P. Wang, "Detecting photographic composites using shadows," in *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*. IEEE, 2009, pp. 1042–1045.
- [37] W. Zhang, X. Cao, Y. Qu, Y. Hou, H. Zhao, and C. Zhang, "Detecting and extracting the photo composites using planar homography and graph cut," *IEEE transactions on information forensics and security*, vol. 5, no. 3, pp. 544–555, 2010.
- [38] X. Zhao, J. Li, S. Li, and S. Wang, "Detecting digital image splicing in chroma spaces," in *International Workshop on Digital Watermarking*. Springer, 2010, pp. 12–22.
- [39] Q. Liu, X. Cao, C. Deng, and X. Guo, "Identifying image composites through shadow matte consistency," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1111–1122, 2011.
- [40] N. B. A. Warif, M. Y. I. Idris, A. W. A. Wahab, and R. Salleh, "An evaluation of error level analysis in image forensics," in *System Engineering and Technology (ICSET), 2015 5th IEEE International Conference on*. IEEE, 2015, pp. 23–28.
- [41] T. S. Gunawan, S. A. M. Hanafiah, M. Kartiwi, N. Ismail, N. F. Za'bah, and A. N. Nordin, "Development of photo forensics algorithm by detecting photoshop manipulation using error level analysis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 7, no. 1, pp. 131–137, 2017.
- [42] D. C. Jeronymo, Y. C. C. Borges, and L. dos Santos Coelho, "Image forgery detection by semi-automatic wavelet soft-thresholding with error level analysis," *Expert Systems with Applications*, vol. 85, pp. 348–356, 2017.
- [43] A. Alahmadi and M. Hussain, "Passive Detection of Image Forgery Using DCT and Local Binary Pattern," *Signal, Image and Video Processing*, vol. 11, no. 1, pp. 81–88, January 2017.
- [44] Y. Q. Shi, C. Chen, and W. Chen, "A Natural Image Model Approach to Splicing Detection," in *Proceedings of the 9th workshop on Multimedia & security*, Dallas, Texas, September 2007, pp. 51–62.
- [45] X. Zhao and J. Li, "Detecting Digital Image Splicing in Chroma Spaces," in *Digital Watermarking*, vol. 6526. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 12–22.
- [46] J. R. Corripio, D. M. Arenas González, A. L. Sandoval Orozco, L. J. García Villalba, J. Hernandez-Castro, and S. J. Gibson, "Source smartphone identification using sensor pattern noise and wavelet transform," 2013.
- [47] A. L. Sandoval Orozco, D. M. Arenas González, J. R. Corripio, L. J. Villalba, and J. C. Hernandez-Castro, "Source identification for mobile devices, based on wavelet transforms combined with sensor imperfections," *Computing*, vol. 96, no. 9, pp. 829–841, 2014.
- [48] A. L. Sandoval Orozco, D. M. Arenas González, J. Rosales Corripio, L. J. García Villalba, and J. C. Hernandez-Castro, "Source Identification for Mobile Devices, Based on Wavelet Transforms Combined with Sensor Imperfections," *Computing*, vol. 96, no. 9, pp. 829–841, September 2014.
- [49] A. Bharati, R. Singh, M. Vatsa, and K. W. Bowyer, "Detecting Facial Retouching Using Supervised Deep Learning," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1903–1913, September 2016.
- [50] C. Chen, A. Dantcheva, and A. Ross, "Automatic Facial Makeup Detection with Application in Face Recognition," in *Proceedings of the International Conference on Biometrics (ICB)*, Madrid, Spain, June 2013, pp. 1–8.
- [51] A. Dantcheva, C. Chen, and A. Ross, "Can Facial Cosmetics Affect the Matching Accuracy of Face Recognition Systems?" in *Proceedings of the IEEE 5th International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. Washington DC, USA: IEEE, September 2012, pp. 391–398.
- [52] N. Kose, L. Apvrille, and J. L. Dugelay, "Facial Makeup Detection Technique Based on Texture and Shape Analysis," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1, Ljubljana, Slovenia, May 2015, pp. 1–7.
- [53] Q. Fu, X. Zhou, C. Wang, and B. Jiang, "Mathematical relation between APBT-based and DCT-based JPEG image compression schemes," *Journal of Communications*, vol. 11, pp. 84–92, January 2016.
- [54] I. IFS-TC, "IFS-TC Image Forensics Challenge," <http://ifc.recod.ic.unicamp.br/>, January 2014.
- [55] Z. Zhang, D. Wang, C. Wang, and X. Zhou, "Detecting Copy-move Forgeries in Images Based on DCT and Main Transfer Vectors," *KSII Transactions on Internet and Information Systems*, vol. 11, pp. 4567–4587, September 2017.

Esteban Alejandro Armas Vega received his Computer Science degree in 2009 at the Polytechnic Institute "José Antonio Echeverría" in Havana (Cuba) and a M.Sc. degree in Computer Science in 2016 from the Universidad Complutense de Madrid (Spain). He is currently a Ph.D. student in the Department of Software Engineering and Artificial Intelligence of the Faculty of Computer Science and Engineering at the Universidad Complutense de Madrid (UCM) and Member of the Complutense Research Group GASS (Group of Analysis, Security and Systems, <http://gass.ucm.es>). His research interests include computer networks and computer security.

Ana Lucila Sandoval Orozco was born in Chivolo, Magdalena, Colombia in 1976. She received a Computer Science Engineering degree from the Universidad Autónoma del Caribe (Colombia) in 2001. She holds a Specialization Course in Computer Networks (2006) from the Universidad del Norte (Colombia), and holds a M.Sc. in Research in Computer Science (2009) and a Ph.D. in Computer Science (2014), both from the Universidad Complutense de Madrid (Spain). She is currently a postdoctoral researcher and member of the Research Group GASS (Group of Analysis, Security and Systems, <http://gass.ucm.es>) at Universidad Complutense de Madrid (Spain). Her main research interests are coding theory, information security and its applications.

Luis Javier García Villalba received a Telecommunication Engineering degree from the Universidad de Málaga (Spain) in 1993 and holds a Ph.D. in Computer Science (1999) from the Universidad Politécnica de Madrid (Spain). Visiting Scholar at COSIC (Computer Security and Industrial Cryptography, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, Belgium) in 2000 and Visiting Scientist at IBM Research Division (IBM Almaden Research Center, San Jose, CA, USA) in 2001 and 2002, he is currently Associate Professor of the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM) and Head of Complutense Research Group GASS (Group of Analysis, Security and Systems) which is located in the Faculty of Computer Science and Engineering at the UCM Campus.

His professional experience includes the management of both national and international research projects and both public (Spanish Ministry of R&D, Spanish Ministry of Defence, Horizon 2020 - European Commission, . . .) and private financing (Hitachi, IBM, Nokia, Safelayer Secure Communications, TB Solutions Security, . . .). Author or co-author of numerous international publications is editor or guest editor of numerous journals such as Entropy MPDI, Future Generation Computer Systems (FGCS), Future Internet MDPI, IEEE Latin America Transactions, IET Communications (IET-COM), IET Networks (IET-NET), IET Wireless Sensor Systems (IET-WSS), International Journal of Ad Hoc and Ubiquitous Computing (IJAHUC), International Journal of Multimedia and Ubiquitous Engineering (IJMUE), Journal of Supercomputing, Sensors MDPI, etc.

Detección de Post-Procesamiento en Vídeos Digitales de Dispositivos Móviles mediante el Análisis de la Estructura del Contenedor Multimedia

Carlos Quinto Huamán, Daniel Povedano Álvarez, Ana Lucila Sandoval Orozco, and Luis Javier García Villalba, *Member, IEEE*

Resumen—Technological innovations have increased the usefulness of mobile devices and a mobile phone is the most useful tool for any daily task since, in addition to providing access to a wide variety of digital content (instant messaging applications, social networks, services, tv, and so on), it allows the generation of multimedia content (videos and images) taking advantage of the advantages offered by integrated cameras. It is increasingly common for users of these devices to share these videos on different platforms to express themselves, without knowing that this content is exposed to any manipulation, compromising its authenticity and integrity. Likewise, videos shared on social networks and through instant messaging applications go through filtering and compression processes to reduce their size, facilitate their transfer, and optimize storage on their platforms. The result of these transformations leaves a distinctive pattern in the multimedia content of the social network used. This work presents a forensic method to identify the characteristic fingerprints that each social network and instant messaging application leaves in the MOV and MP4 videos shared across their platforms. This method is based on the extraction and analysis of the structure of multimedia containers, and the use of supervised machine learning algorithms.

Palabras claves—Container Structure Analysis, Forensics Analysis, Multimedia Container, Social Media Detection, Supervised Classification Techniques.

I. INTRODUCCIÓN

Los teléfonos móviles se han convertido en una herramienta fundamental para las personas, debido que ofrece realizar múltiples actividades con un único dispositivo, como acceder a Internet, enviar correos electrónicos, usar la cámara integrada para capturar contenido multimedia, usar aplicaciones que solucionan operaciones que antes demandaba mucho más tiempo (transacciones bancarias, solicitar citas médicas, compras online, etc). Esto lo convierte en uno de los dispositivos más demandados en los últimos años y se prevé un crecimiento en los próximos. Según la investigación de Cisco [1], en 2022, el tráfico IP global alcanzará los 396 Exabytes mensuales (4,8 Zettabytes anuales), los usuarios de Internet aumentarán a 4.800 millones de los 3.400 millones del 2017 y habrá 28.500 millones de conexiones de dispositivos personales fijos y móviles de los 18.000 millones del

2017. Asimismo, el tráfico de vídeo IP representará un 82 % del tráfico global IP, comparado con el 75 % que se alcanzó en 2017.

Un reciente estudio de Hootsuite [2] indica que actualmente un usuario medio de Internet pasa más de 6 horas y media en línea cada día, lo que significa que la comunidad digital del mundo pasará más de 1.200 millones de años utilizando Internet en 2019. También señalan que este año existe alrededor de 3.484 millones de usuarios activos de redes sociales, que representa un 9 % más que en 2018. De estos, 3.256 millones acceden a estas a través de los teléfonos móviles. También indican que la red social Facebook es la más popular con 2.120 millones de usuarios activos mensualmente en todo el mundo, seguida por Youtube con 1.900 millones y por WhatsApp con 1.500 millones. En cuanto aplicaciones de mensajería instantánea WhatsApp es la preferida por 133 países del mundo y Facebook messenger en 75 países.

Es evidente que la tecnología provee de múltiples beneficios a la cotidianidad de la sociedad actual, sin embargo, estos beneficios se pueden convertir en un puente o conexión para que personas malintencionadas aprovechen los recursos que se encuentran sin la protección debida y realicen algún tipo de fraude o falsificación. Los vídeos capturados por dispositivos móviles no están exentos a este tipo de amenazas, porque es común compartir este tipo de ficheros por redes sociales, dispositivos de almacenamiento, incluso por la pérdida del dispositivo móvil. Por lo anterior, los vídeos están cada vez más propensos a ser manipulados y ser presentados como pruebas digitales en ámbitos procesales para evadir responsabilidades sobre acciones delictivas como pornografía infantil, tráfico de personas, etc. En este sentido, es necesario investigar sobre diferentes métodos para verificar la autenticidad e integridad de un vídeo. En esta investigación, se presenta un método para la detección de post-procesamiento de vídeos compartidos por las plataformas de redes sociales y aplicaciones de mensajería instantánea basado en el análisis de la estructura del contenedor multimedia combinado con el uso de algoritmos de aprendizaje supervisado.

Este trabajo está estructurado en 6 secciones, siendo la primera la presente introducción. En la Sección II se describe brevemente algunos conceptos sobre contenedores multimedia. La Sección III estudia los trabajos relacionados a la identificación de la fuente de vídeos utilizando la estructura de contenedores multimedia. En la Sección IV se presenta el método propuesto. En la Sección V se describen los

C. Quinto Huamán, D. Povedano Álvarez, A. L. Sandoval Orozco and L. J. García Villalba. Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Ingeniería del Software e Inteligencia Artificial, Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, España. e-mail: {cquinto, dpovedano}@ucm.es, {asandoval, javiergv}@fdi.ucm.es.

experimentos y resultados. Por último en la Sección VI se presentan las conclusiones del presente trabajo.

II. CONTENEDORES MULTIMEDIA

El contenedor multimedia es un formato de archivo que contiene varios tipos de datos como el flujo de datos de vídeo y audio, previamente comprimidos por un determinado códec, metadatos, además de otros elementos como subtítulos. Para encapsular o envolver estos datos los fabricantes siguen un formato preestablecido en las especificaciones técnicas de estándares internacionales. En [3] [4] indican que un contenedor multimedia está constituido por una pista de vídeo y una de audio. Ambas pistas son comprimidas por un códec propietario de cada fabricante.

En [5] se indica que la compresión del vídeo se consigue mediante la eliminación de la redundancia temporal que existe entre los fotogramas que componen una secuencia, de esta forma sólo se obtienen componentes necesarios para una reproducción fiel de los datos. Existen diferentes estándares de compresión de vídeos, pero en la actualidad los más usados por los dispositivos móviles son dos: 1) H264/AVC o MPEG-4 Parte 10 [6] y 2) H265/HEVC o MPEG-H Parte 2 [7], ambos desarrollados por el ITU-T *International Telecommunication Union* (ITU) y ISO/IEC *International Organization for Standardization* (ISO), *International Electrotechnical Commission* (IEC). Los tres contenedores multimedia más usados en la actualidad son: 1) Contenedor MP4 del estándar *Moving Picture Experts Group* (MPEG)-4 parte 14. 2) Contenedor 3GP que cumple las especificaciones MPEG-4 parte 14. Los fabricantes cuyo sistema operativo es Android [8], utilizan principalmente los dos primeros contenedores. 3) Contenedor MOV del estándar *QuickTime*, desarrollado por *Apple* [9]. Este último contenedor es usado por los dispositivos cuyo sistema operativo es iOS [10]. No obstante, existen otros contenedores multimedia como *Audio Interleave de Microsoft* (AVI) [11] y *Matroska* (MKV) [12] que son usados por dispositivos más específicos. En este trabajo se utilizan vídeos con contenedores multimedia MOV y MP4, debido a su alta presencia en los principales fabricantes de dispositivos móviles. Los contenedores multimedia de vídeos están estructurados por átomos o cajas que sirven de soporte fundamental para el almacenamiento de toda la información necesaria para la reproducción del vídeo [13]. Un átomo contiene una cabecera, seguido de los datos del propio átomo y estos átomos están organizados de forma jerárquica, es decir un átomo puede contener otros átomos y así sucesivamente, pero cuando un átomo no contiene otros átomos se le denomina átomo hoja o *leaf atom* [6] [9]. La cabecera de un átomo contiene generalmente los campos *size* y *type*, excepcionalmente el campo *extended size*. Adicionalmente los átomos tienen uno o más etiquetas con sus respectivos valores. El campo *type* se traduce en un entero sin signo de 32 bits, interpretado como un código *American Standard Code for Information Interchange* (ASCII) de cuatro caracteres. El campo *size* se traduce en un entero de 32 bits, interpretado como un código de cuatro caracteres. El tamaño real de un átomo no puede ser menor de 8 bytes dado que siempre contienen los campos *type* y *size*. En este sentido, el átomo se representa como: */moov/*; el átomo que contiene otro átomo hijo como: */moov/mvhd/*; el

átomo con etiqueta y valor como: */moov/mvhd/version, value: 0)*). Este conjunto de características se denomina estructura del contenedor multimedia. Cada red social y aplicación de mensajería instantánea incrusta una estructura propia al momento de compartir los vídeos; esto se debe a que ejecutan un proceso de re-compresión sobre los vídeos.

III. TRABAJOS RELACIONADOS

Los últimos años el aprendizaje automático o *Machine Learning* (ML) ha contribuido notablemente a la resolución de diversos tipos de tareas, como la gestión de riesgos financieros, detección precoz del cáncer, seguridad de datos, la educación, campañas electorales, ciencia forense, etc. Según [14] ML es un método de análisis de datos que automatiza la construcción de modelos analíticos y está basado a cuatro tipos de aprendizaje: supervisado, semi-supervisado, no supervisado y de refuerzo.

En este contexto, la literatura nos provee trabajos relacionados a la detección de la fuente de vídeos utilizando técnicas de ML, pero en su gran mayoría están orientados a explotar características tradicionales como el ruido del sensor [15] [16] [17] [18] [19] [20] [21], existiendo pocas investigaciones que utilizan la estructura de contenedores multimedia para estos fines.

No obstante, este número reducido de trabajos que explotan la estructura de contenedores multimedia con fines de detección de la fuente, se centran en vídeos con formatos AVI, siendo muy limitado para contenedores MOV y MP4. Asimismo, existen investigaciones que analizan la estructura del contenedor multimedia de un número reducido de redes sociales y aplicaciones de mensajería instantánea.

En [18], se realiza un estudio de los tipos características que son objeto de análisis forense en dispositivos móviles. El problema fundamental de este enfoque es que los diferentes modelos de las cámaras digitales usan componentes de un número reducido de fabricantes y que los algoritmos que usan para la generación de las imágenes y vídeos también son muy similares entre modelos de la misma marca.

En [22] se realiza una comparación minuciosa de los principales grupos de técnicas de identificación de la fuente de adquisición. Estas se dividen en cinco grupos y están basadas en: metadatos, características de la imagen, defectos de la matriz *Color Filter Array* (CFA) e interpolación cromática, imperfecciones del sensor y las transformadas wavelet.

En [23] los autores implementan una técnica para verificar la integridad de vídeos con formato AVI, generados por grabadores de datos de eventos (*Video Event Data Recorders* (VEDRS)). Realizaron el análisis de la estructura de 296 vídeos originales que posteriormente fueron editados por 5 programas de edición. Los resultados del análisis demostraron que los editores cambian notablemente la estructura y los valores de los metadatos con respecto los originales. Cada programa de edición incrusta una estructura específica que ayuda al analista forense a detectar si un vídeo ha sufrido algún tipo de manipulación.

En [24] se realizó un análisis de las estructuras de vídeos con formato AVI y MP4, agrupados en 19 modelos de cámaras digitales, 14 modelos de teléfonos móviles y 6 programas de edición. Después de analizar los vídeos originales los

autores determinaron que las estructuras de cada tipo de contenedor no está estrictamente definida como se especifica en los estándares. Se encontraron diferencias considerables entre vídeos generados por dichos dispositivos. Asimismo, los vídeos AVI después de ser manipulados con los programas de edición, cambiaron la estructura interna incluyendo los valores de los metadatos, características esenciales para saber el origen de los vídeos.

En [25], los autores implementaron un método no supervisado para verificar la integridad de vídeos basado en la disimilitud entre vídeos originales y editados. Asimismo, desarrollaron un método para identificar la fuente de adquisición del vídeo mediante el análisis de los contenedores. Para lograr este objetivo utilizaron la librería *MP4Parser*[26], obteniendo ficheros de lenguaje de marcado extensible (*Extensible Markup Language* (XML)) para un posterior análisis, logrando buenos resultados en sus experimentaciones, sosteniendo que la solución utiliza un mínimo de recurso computacional a comparación de otras alternativas.

IV. MÉTODO PROPUESTO

En este trabajo se propone un método forense para la identificación de redes sociales y aplicaciones de mensajería instantánea en vídeos con formato MOV y MP4, utilizando como característica principal la estructura de los contenedores multimedia. La metodología y procedimientos que se han seguido para evaluar y seleccionar el mejor modelo de ML supervisado se detalla en la Figura 1.

IV-A. Extracción de Átomos

El proceso de extracción de características es una parte fundamental en la aplicación de cualquier técnica de aprendizaje automático. Para este trabajo se ha desarrollado un algoritmo de extracción de átomos, que se encarga de obtener un conjunto de características (átomos, etiquetas, valores y orden de aparición) de los vídeos con formato MOV y MP4, para ser analizadas y posteriormente encontrar un patrón que permita distinguir una red social y aplicación de mensajería instantánea. La implementación del algoritmo está basado al uso de funciones recursivas. Primero, se obtiene byte inicial del átomo; el tamaño del átomo (4 bytes); y el tipo de átomo (4 bytes) que está representado por una cadena de caracteres. Segundo, se extrae el átomo leído y se asigna un orden relativo a cada átomo, con la finalidad de organizar eficientemente las estructuras y evitar eliminar átomos que

poseen similar nombre. Tercero, se localiza y extrae los átomos hijos conjuntamente con sus etiquetas y valores; este proceso se repetirá mientras el átomo contenga otros átomos. En caso que el átomo extraído no es ubicado en la lista de átomos previamente implementados, dicho átomo se registra como átomo desconocido. Una vez finalizado el proceso, se almacena la información en un DataFrame.

IV-B. Pre-Procesamiento

Las características que provienen de la extracción de átomos están estructuradas en un DataFrame con los siguientes campos: *Path-file-name*, que es la ruta del vídeo seguido por el nombre; *Path-origin*, es la ruta del vídeo; *Class-label*, es la identificación de cada red social y aplicación de mensajería; *File-name*, es el nombre del vídeo; *Marker*, es la marca del dispositivo móvil; *Model*, es el modelo del dispositivo móvil; *PathOrder-tag* es el conjunto de átomos con su respectivo orden relativo, seguido de la etiqueta con su respectivo valor; *ValueReading-orders*, es el orden absoluto de cada átomo, también llamado orden de aparición.

Dado un vídeo (X), contiene un conjunto de átomos ($a_1, \dots, a_n$), representado por */ftyp-1/*. También, estos átomos pueden contener otros átomos y etiquetas: $a=((a_1), w_1 \dots (a_n), w_n)$ representado por */ftyp-1/majorBrands*. Asimismo, estas etiquetas tienen valores: */ftyp-1/majorBrands: @mp42*. En este sentido, el *PathOrder-tag* contiene 2 tipos de características: Secuencia de átomos (*/ftyp-1/*) y secuencia de átomos seguido de las etiquetas (*/ftyp-1/majorBrands*).

Para el presente trabajo se utiliza solo se utiliza los campos *Path-file-name*, *PathOrder-tag*, *Class-label*, y por ello se elimina las columnas que no son utilizadas. Seguidamente se asigna un identificador a cada clase y se agrupa la estructura de cada vídeo, representado por un vector con N *PathOrder-tag* con igual cardinalidad. Esto se lleva a cabo utilizando la presencia o ausencia de los *PathOrder-tag*, asignando la variable binaria 1 en caso de presencia y 0 en caso de ausencia. Después, se obtiene el conjunto de características de entrenamiento $Train_{features}$ y sus respectivas etiquetas. Finalmente, se divide el $Train_{features}$ 2 partes: 80 % (Conjunto de entrenamiento) y 20 % (Conjunto de test).

IV-C. Oversampling y Reducción de Dimensiones

Un problema común en el uso de técnicas de ML es el desbalanceo entre clases.

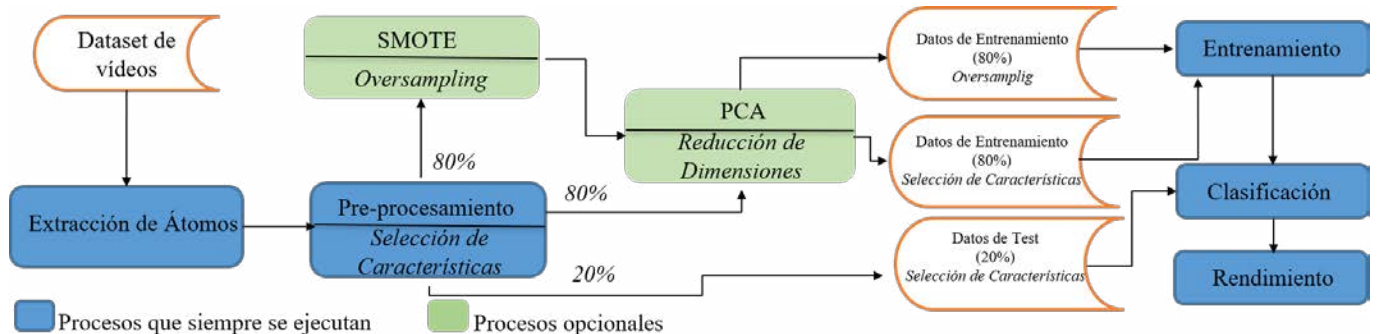


Figura 1: Pipeline del método propuesto

Si bien es cierto, la cantidad de vídeos utilizados para cada clase es el mismo, se ha considerado este paso como opcional para evaluar el funcionamiento del modelo bajo éstas condiciones y así solventar el impacto que tiene sobre la decisión final de los clasificadores.

En este sentido, se hace uso de la técnica muy utilizada en la literatura denominado *Synthetic Minority OverSampling* (SMOTE) [27], que genera o incrementa el número de instancias de las clases minoritarias de forma sintética [28] [29], en este contexto cada instancia es un vector por cada vídeo con N *PathOrder-tag*.

Otro problema frecuente es contar con características menos importantes y que no aportan significativamente a la predicción. A esta problemática se suma un mayor tiempo de procesamiento y mayor uso de recursos. Para mitigar esta dificultad, opcionalmente se hace uso de la técnica de Análisis de Componentes Principales o *Principal Component Analysis* (PCA), que ayuda a identificar y eliminar *PathOrder-tag* irrelevantes para mejorar el rendimiento computacional y entender mejor la evaluación del modelo y sus resultados. Cabe señalar, que a ésta técnica se le antepone la técnica SMOTE a fin de evitar causar un sobreajuste o *Overfitting* al modelo [30].

IV-D. Training

La fase de entrenamiento es fundamental ya que determina el rendimiento de los algoritmos y de esta forma elegir el mejor modelo predictivo para realizar la clasificación. Es común utilizar un buen porcentaje de características, dejando una porción para evaluar el rendimiento y capacidad de generalizar nuevas características que no han sido contemplados en la fase de entrenamiento. En nuestro caso, como se detalló en la Sección IV-B se destina un 80 % de *PathOrder-tag* para el entrenamiento (Conjunto de entrenamiento) y un 20 % para evaluar el modelo (Conjunto de Test).

Una vez definido el porcentaje de los datos de entrenamiento, el siguiente paso es seleccionar los hiper-parámetros de los modelos de ML y esto depende del cada algoritmo que se utilizará. Existen varios procedimientos para obtener hiper-parámetros que ofrezcan un buen rendimiento del algoritmo de clasificación [31]. En este trabajo se utilizó la técnica búsqueda exhaustiva o *grid search* [28]. Para ello, se construye los *pipelines* para cada *ensemble classification algorithm*: *Logistic Regression* (LR), *Support Vector Machine* (SVM), *Decision Tree* (DT), *Linear Discriminant Analysis* (LDA), *Gaussian Naive Bayes* (GNB), *KNeighbors* (KNN), y para cada *Non-ensemble classification algorithm*: *Random Forest* (RF), *Gradient Boosting Classifier* (GBC), *Extreme Gradient Boosting* (XGB) [29] [32]. El conjunto de hiper-parámetros seleccionados para cada uno de los algoritmos propuestos, aplicando la validación cruzada o *cross-validation* con $k=5$ se detalla a continuación:

- **Logistic Regression (LR)**: *clf_C*: 0.1, *clf_penalty*: l2, *clf_solver*: liblinear.
- **Support Vector Machine (SVM)**: *clf_C*: 0.1, *clf_kernel*: linear.
- **Decision Tree (DT)**: *clf_min_samples_split*: 10, *clf_min_samples_leaf*: 1, *clf_max_depth*: 18.
- **Linear Discriminant Analysis (LDA)**: *clf_solver*: svd

- **Gaussian Naive Bayes (GNB)**: *clf_var_smoothing*: 1e-08.
- **KNeighbors (KNN)**: *clf_weights*: distance, *clf_n_neighbors*: 15, *clf_metric*: cosine.
- **Random Forest (RF)**: *clf_n_estimators*: 1200, *clf_min_samples_split*: 10, *clf_min_samples_leaf*: 4, *clf_max_features*: sqrt, *clf_max_depth*: 10, *clf_bootstrap*: True.
- **Gradient Boosting Classifier (GBC)**: *clf_subsample*: 1.0, *clf_n_estimators*: 100, *clf_min_samples_split*: 0.9, *clf_min_samples_leaf*: 0.2090.
- **Extreme Gradient Boosting (XGB)**: *clf_subsample*: 0.6, *clf_min_child_weight*: 5, *clf_max_depth*: 4, *clf_gamma*: 2, *clf_colsample_bytree*: 0.8.

IV-E. Testing

En esta fase se hace uso del conjunto de test, es decir el 20 % de *PathOrder-tag* separado en la Sección IV-B, con el objetivo de evaluar la capacidad de clasificación del métodos utilizados. Asimismo, se valida si los métodos se adecuan a un entorno real. El resultado de la evaluación será la elección del modelo que mejor se comporta con el tipo de características que se provee. Para ello, se utiliza 4 métricas de rendimiento que se presentan en a siguiente sección.

IV-F. Métricas de Rendimiento

Como se mencionó en la Sección III, en la actualidad, no existe investigaciones relacionadas a la clasificación supervisada de redes sociales y aplicaciones de mensajería instantánea que utilicen la estructura del contenedor multimedia, y que sirvan como referencia para elegir las mejores métricas. En este contexto, se consideró utilizar 4 métricas para la evaluación del rendimiento de soluciones de identificación de la fuente de adquisición en vídeos (*Accuracy*, *F1-Score-micro*, *F1-Score-macro*, *Log Loss*) [29]. Considerando que se desea clasificar cada muestra (vector con N variables) con su respectiva clase, se denomina verdaderos positivos (*TP*) a las muestras que se clasificaron como positivos correctamente, verdaderos negativos (*TN*) a las muestras que se clasificaron como negativos correctamente, falsos positivos (*FP*) a la predicción no clasificada en la clase de interés, falsos negativos (*FN*) a la predicción no clasificada correctamente en la clase de no interés. Previamente a la evaluación de las métricas propuestas, se evaluó dos métricas de soporte: *Recall*, que es la tasa de verdaderos positivos, es decir nos indica la capacidad del método para clasificar las muestras en la clase correcta; *Precision*, nos indica la proporción de muestras que nuestro método ha clasificado en una determinada clase. En los siguientes párrafos se describen brevemente las métricas utilizadas:

- **Accuracy**. Esta métrica es la más simple de calcular y es el porcentaje total de elementos clasificados correctamente, se define como:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- **F1-Score**. Es la media armónica de *Precision* y *Recall*. Mientras el valor se acerque más a 1, el rendimiento del sistema de clasificación será más óptimo [33]. En

este trabajo, ésta métrica se desglosa en 2, de acuerdo al tipo de configuración del cálculo de la media: 1) F1-macro (Average:macro) que calcula la métrica independientemente para cada clase y luego toma el promedio, es decir, trata a todas las clases por igual; 2) F1-micro (Average:micro) agrega las contribuciones de todas las clases para calcular la métrica promedio. En un sistema de clasificación multiclase, es preferible utilizar esta métrica si se sospecha que puede haber un desequilibrio de clase, es decir puede tener muchos más vídeos de una clase que de otras clases. La métrica se define como:

$$F1-Score = 2 * \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

- **Log-loss.** Mide el rendimiento de un modelo de clasificación donde la entrada de predicción es un valor de probabilidad entre 0 y 1. El objetivo de nuestros modelos es minimizar este valor. Un modelo perfecto tendría una pérdida logarítmica de 0, se define como:

$$Log-loss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}) \quad (3)$$

Donde: N es el número de muestras o instancias, M es el número de etiquetas posibles, y_{ij} toma el valor de 1 para la clase correcta y 0 para otras clases, p_{ij} es la probabilidad asignada para esa clase.

Después de haber detallado el método propuesto, se realiza la evaluación de cada uno de los algoritmos seleccionados en la Sección IV-D, con la finalidad de seleccionar un modelo robusto y eficiente. En la Tabla I se observa el rendimiento de los algoritmos para la identificación de redes sociales y aplicaciones de mensajería instantánea utilizando PCA y descartando SMOTE. Se descartó el uso de SMOTE, debido a que los resultados obtenidos en experimentaciones previas son similares a la Tabla I, demostrando que el dataset de vídeos está balanceado.

Tabla I: Evaluación del rendimiento de los modelos candidatos

Modelo	Accuracy	F1-micro	F1-macro	Log loss
LR	0.975891	0.975891	0.972057	0.063861
RF	0.975891	0.975891	0.972057	0.063479
DT	0.975552	0.975552	0.971684	0.078106
SVM	0.975891	0.975891	0.972057	0.074281
KNN	0.975891	0.975891	0.972057	0.067543
LDA	0.927674	0.927674	0.910590	0.760697
GNB	0.975891	0.975891	0.972057	0.832684
GBC	0.707640	0.707640	0.626514	0.656345
XGB	0.975891	0.975891	0.972057	0.075127

En general, se observa un rendimiento muy alto de todos los algoritmos, excepto GBC, que alcanza un accuracy y F1-micro de 0.707640 y F1-macro de 0.626514. No obstante, LDA, GNB y GBC alcanzan valores log-loss cercanos a 1, lo que significa que el rendimiento en la clasificación de estos algoritmos no son tan perfectos. El modelo con mejor rendimiento es RF, alcanzando excelentes resultados en las 4 métricas. En este sentido, el algoritmo RF, será utilizado en las experimentaciones del presente trabajo.

V. EXPERIMENTOS Y RESULTADOS

Para evaluar el método propuesto se han realizado dos grupos de experimentos: un análisis de la estructura de

contenedores multimedia y, posteriormente, la identificación de redes sociales y aplicaciones de mensajería instantánea utilizando el modelo de ML seleccionado.

V-A. Descripción del Dataset

La preparación del conjunto de datos es un proceso muy importante para este trabajo ya que es un factor que determina la calidad de resultados que se desea alcanzar. En la literatura, existen muy pocos conjuntos de datos de vídeos generados por dispositivos móviles que hayan sido compartidos por redes sociales y aplicaciones de mensajería instantánea. Tener un conjunto de datos de vídeos organizado, robusto, balanceado y sobre todo actualizado es todo un reto. Por ese motivo, se ha generado un conjunto de datos heterogéneo y suficientemente grande, para evaluar el método propuesto en escenarios reales y así obtener resultados efectivos. Las Tablas II y III conforman el dataset generado para los experimentos. En la Tabla II se presenta el conjunto de datos que está compuesto por 270 vídeos originales, 10 vídeos por cada modelo. Los vídeos fueron capturados con las opciones de las cámaras por defecto (p. ej. registro de ubicación desactivado). Posteriormente, los vídeos de la Tabla II fueron compartidos a través de 8 redes sociales (2160 vídeos resultantes) y 3 aplicaciones de mensajería instantánea (810 vídeos resultantes), haciendo un total de 2970 vídeos que serán analizados. En la Tabla III se detalla las redes sociales y aplicaciones de mensajería instantánea que se utilizaron con sus respectivas configuraciones de subida y descarga de los vídeos.

Tabla II: Características de los vídeos originales del dataset

Marca	Modelo	Id Modelo	S.O.	Resolución	Codec
Apple	Ipad 2	D01	iOS 9.3.5	1280 x 720p	H.264
	Ipad Air	D02	iOS 11.3	1920 x 1080p	H.264
	Iphone 5	D03	iOS 7.0.4	1920 x 1080p	H.264
	Iphone 5S	D04	iOS 9.2	1920 x 1080p	H.264
	Iphone 6	D05	iOS 8.4	1920 x 1080p	H.264
	Iphone 7	D06	iOS 11.2.6	3840 x 2160p	H.264
	Iphone 8 Plus	D07	iOS 11.2.5	1920 x 1080p	H.265
	Iphone X	D08	iOS 11.4.1	1920 x 1080p	H.264
	Iphone XS Max	D09	iOS 12.1.0	3840 x 2160p	H.264
Huawei	Ascend	D10	Android	1280 x 720p	H.264
	P9	D11	Android	1920 x 1080p	H.264
	P10	D12	Android	1920 x 1080p	H.264
LG	Nexus 5	D13	Android	1920 x 1080p	H.264
	G6	D14	Android	640x480p	H.264
Microsfot	Lumia 640 LTE	D15	Windows Phone	1920 x 1080p	H.264
Motorola	Moto G2	D16	Android	1280 x 720p	H.264
	Nexus 6	D17	Android	1920 x 1080p	H.264
One Plus	A0001	D18	Android	1920 x 1080p	H.264
Samsung	Galaxy A6	D19	Android	1920 x 1080p	H.264
	Galaxy S5	D20	Android	3840 x 2160p	H.264
	Galaxy S7	D21	Android	1920 x 1080p	H.264
	Galaxy S9 Plus	D22	Android	3840 x 2160p	H.265
	Galaxy J5 2016	D23	Android	1920 x 1080p	H.264
	Galaxy Tab A	D24	Android	1280 x 720p	H.264
Xiaomi	Mi3	D25	Android	1280 x 720p	H.264
	Redmi Note 5	D26	Android	1920 x 1080p	H.264
	PocoPhone	D27	Android	3840 x 2160p	H.264

V-B. Análisis de Contenedores Multimedia

Este análisis se realiza para observar el comportamiento de las redes sociales y aplicaciones de mensajería instantánea a la hora de subir y descargar vídeos a través de sus plataformas. Asimismo, conocer de manera general cual es la estructura que inserta cada herramienta a los contenedores multimedia. En este sentido, se realiza la comparación de la estructura del

Tabla III: Configuración del proceso de generación del dataset de redes sociales y aplicaciones de mensajería instantánea

Red Social	Versión	Proceso de Subida	Proceso de Descarga
Facebook HD	Website	Max 4gb, 240min	Firefox (Inspect element)
Facebook SD	Website	Max 4gb, 240min	Firefox (Inspect element)
Youtube	Website	Max 128gb, 12hrs	Youtube studio beta
Flickr	Website	Max 1gb	Website(save as)
Linkedin	Website	Max 6gb, min 75kb	Website(save as)
Instagram	Website	Max 10 min, ratio 9:16	Firefox (Inspect element)
Twitter	Website	Max 500mb, 2.20 min	Twitervideodownloader
Tumblr	Website	Max 100mb	Firefox (Inspect element)
Aplicación	Versión	Características	
Facebook Msn	255.0.0.13.113	-	-
WhatsApp	2.19.20	-	-
Telegram	5.7.1	Max 1.5GB	-

contenedor de un vídeo original, generado por un teléfono móvil de marca Apple y modelo Ipad Air, con la estructura del mismo vídeo luego de haber sido compartido a través de Facebook HD, Youtube, WhatsApp, Linkedin, Telegram. Ver Figura 2.

En vídeo original de marca Apple y modelo Ipad Air, tiene 4 átomos raíz (*ftyp-1, wide-2, mdat-3, moov-4*). Los valores de las etiquetas *@majorBrands* y *@compatibleBrand* son *qt* y *qt* respectivamente. La ruta para obtener el valor de la etiqueta *@timeScale* es */ftyp-1/wide-2/mdat-3/moov-4/mvhd-1/@timeScale:600*. El átomo *moov-4* contiene 4 átomos *trak*, 1 para la pista de vídeo, 1 para la pista de audio y 2 exclusivamente para metadatos. La estructura del contenedor esta compuesto por 651 características.

En vídeo de Facebook HD, tiene 4 átomos raíz (*ftyp-1, moov-2, free-3, mdat-4*). Los valores de las etiquetas *@majorBrands* y *@compatibleBrand* son *isom* y *isomiso2avc1mp41* respectivamente. La ruta para obtener el valor de la etiqueta *@timeScale* es */ftyp-1/moov-2/mvhd-1/@timeScale:1000*. El átomo *moov-2* contiene 2 átomos *trak*, 1 para la pista de vídeo, 1 para la pista de audio. La estructura del contenedor esta compuesto por 267 características.

En vídeo de Youtube, tiene 3 átomos raíz (*ftyp-1, moov-2, mdat-3*). Los valores de las etiquetas *@majorBrands* y *@compatibleBrand* son *mp42* y *isommp42* respectivamente. La ruta para obtener el valor de la etiqueta *@timeScale* es */ftyp-1/moov-2/mvhd-1/@timeScale:1000*. El átomo *moov-2* contiene 2 átomos *trak*, 1 para la pista de vídeo, 1 para la pista de audio. La estructura del contenedor está compuesto por 250 características.

En vídeo de WhatsApp, tiene 4 átomos raíz (*ftyp-1, beam-2, moov-3, mdat-4*). Los valores de las etiquetas *@majorBrands* y *@compatibleBrand* son *mp42* y *mp41mp42isom* respectiva-

mente. La ruta para obtener el valor de la etiqueta *@timeScale* es */ftyp-1/beam-2/moov-3/mvhd-1/@timeScale:44100*. El átomo *moov-3* contiene 2 átomos *trak*, 1 para la pista de vídeo, 1 para la pista de audio. La estructura del contenedor esta compuesto por 237 características.

En vídeo de Linkedin, tiene 4 átomos raíz (*ftyp-1, moov-2, free-3, mdat-4*). Los valores de las etiquetas *@majorBrands* y *@compatibleBrand* son *isom* y *isomiso2avc1mp41* respectivamente. La ruta para obtener el valor de la etiqueta *@timeScale* es */ftyp-1/moov-2/mvhd-1/@timeScale:1000*. El átomo *moov-2* contiene 2 átomos *trak*, 1 para la pista de vídeo, 1 para la pista de audio. La estructura del contenedor está compuesto por 258 características.

En vídeo de Telegram, tiene 3 átomos raíz (*ftyp-1, mdat-2, moov-3*). Los valores de las etiquetas *@majorBrands* y *@compatibleBrand* son *mp42* y *mp41mp42isom* respectivamente. La ruta para obtener el valor de la etiqueta *@timeScale* es */ftyp-1/mdat-2/moov-3/mvhd-1/@timeScale:44100*. El átomo *moov-3* contiene 2 átomos *trak*, 1 para la pista de vídeo, 1 para la pista de audio. La estructura del contenedor está compuesto por 251 características.

Cabe señalar que después de analizar los vídeos de la red social Flickr, se ha confirmado que la plataforma no realiza ningún tipo de re-compresión y tampoco cambia la estructura del contenedor multimedia, manteniendo las mismas características que el vídeo original. En resumen, este análisis ha permitido demostrar demostrar que cada red social y aplicación de mensajería instantánea inserta una estructura diferente. Estas diferencias se derivan del orden de aparición de los átomos y valores asignados cada etiquetas.

V-C. Identificación de Redes Sociales y Aplicaciones de Mensajería Instantánea

En este experimento se utilizaron los 2970 videos compartidos en 8 redes sociales y 3 aplicaciones de mensajería instantanea, repartidos en 2376 videos para train y 594 para test. En la Tabla IV se observa que WhatsApp, Facebook HD, Facebook Msn, Telegram, Youtube, Flickr, Linkedin, Twitter y Tumblr se clasifican al 100 %. Sin embargo la red social Facebook SD alcanza un 0.98 % de acierto, confundándose un 0.02 % con Tumblr. Finalmente, Instagram alcanza un 0.69 % de acierto, confundándose un 0.31 % con Facebook Msn. En la Figura 3, se observa un resumen del factor de impacto de los *PathOrder-tag* o características que tienen sobre el modelo a la hora de clasificar una clase. Los *PathOrder-tag* que contienen el atomo *wide-2* tienen mayor impacto para definir que la clase sea Flickr (vídeos originales).

Tabla IV: Matriz de confusión para redes sociales y aplicaciones de mensajería instantánea

Social media	WhatsApp	FacebookHD	FacebookSD	FacebookMsn	Telegram	Youtube	Flickr	Linkedin	Instagram	Twitter	Tumblr
WhatsApp	1	-	-	-	-	-	-	-	-	-	-
FacebookHD	-	1	-	-	-	-	-	-	-	-	-
FacebookSD	-	-	0.98	-	-	-	-	-	-	-	0.02
FacebookMsn	-	-	-	1	-	-	-	-	-	-	-
Telegram	-	-	-	-	1	-	-	-	-	-	-
Youtube	-	-	-	-	-	1	-	-	-	-	-
Flickr	-	-	-	-	-	-	1	-	-	-	-
Linkedin	-	-	-	-	-	-	-	1	-	-	-
Instagram	-	-	-	0.31	-	-	-	-	0.69	-	-
Twitter	-	-	-	-	-	-	-	-	-	1	-
Tumblr	-	-	-	-	-	-	-	-	-	-	1

iPhone Ipad Air

```

ftyp-1
├─ @majorBrands: qt
├─ @minorVersion: 0
├─ @compatibleBrand :
  qt
wide-2
mdat-3
moov-4
├─ mvhd-1
│   └─ @version: 0
│   └─ @flags: 0
│   └─ @reserved: 0
│   └─ @timeScale: 600
│   └─ @posterTime: 0
│   └─ @currentTime: 0
│   └─ @nextTrackId: 5
│   └─ @previewTime: 0
│   ...
├─ trak-1
│   └─ tkhd-1
│       └─ @version: 0
│       └─ @flags: 0
│       ...
├─ trak-2 ...
├─ trak-3 ...
├─ trak-4 ...
Total structure = 651

```

Facebook HD

```

ftyp-1
├─ @majorBrands: isom
├─ @minorVersion: 512
├─ @compatibleBrand:
  isomiso2avc1mp41
moov-2
├─ mvhd-1
│   └─ version: 0
│   └─ @flags: 0
│   └─ @reserved: 0
│   └─ @timeScale: 1000
│   └─ @posterTime: 0
│   └─ @currentTime: 0
│   └─ @nextTrackId: 3
│   └─ @previewTime: 0
│   ...
├─ trak-1
│   └─ tkhd-1
│       └─ @version: 0
│       └─ @flags: 3
│       ...
├─ trak-2
├─ free-3
├─ mdat-4
Total structure = 267

```

Youtube

```

ftyp-1
├─ @majorBrands: mp42
├─ @minorVersion: 0
├─ @compatibleBrand:
  isommp42
moov-2
├─ mvhd-1
│   └─ version: 0
│   └─ @flags: 0
│   └─ @reserved: 0
│   └─ @timeScale: 1000
│   └─ @posterTime: 0
│   └─ @currentTime: 0
│   └─ @nextTrackId: 3
│   └─ @previewTime: 3
│   ...
├─ trak-1
│   └─ tkhd-1
│       └─ @version: 1
│       └─ @flags: 3
│       ...
├─ trak-2
├─ mdat-3
Total structure = 250

```

WhatsApp

```

ftyp-1
├─ @majorBrands: mp42
├─ @minorVersion: 1
├─ @compatibleBrand:
  mp41mp42isom
beam-2
moov-3
├─ mvhd-1
│   └─ version: 0
│   └─ @flags: 0
│   └─ @reserved: 0
│   └─ @timeScale: 44100
│   └─ @posterTime: 0
│   └─ @currentTime: 0
│   └─ @nextTrackId: 3
│   └─ @previewTime: 0
│   ...
├─ trak-1
│   └─ tkhd-1
│       └─ @version: 0
│       └─ @flags: 1
│       ...
├─ trak-2
├─ mdat-4
Total structure = 237

```

Linkedin

```

ftyp-1
├─ @majorBrands: isom
├─ @minorVersion: 512
├─ @compatibleBrand:
  isomiso2avc1mp41
moov-2
├─ mvhd-1
│   └─ version: 0
│   └─ @flags: 0
│   └─ @reserved: 0
│   └─ @timeScale: 1000
│   └─ @posterTime: 0
│   └─ @currentTime: 0
│   └─ @nextTrackId: 3
│   └─ @previewTime: 0
│   ...
├─ trak-1
│   └─ tkhd-1
│       └─ @version: 0
│       └─ @flags: 3
│       ...
├─ trak-2
├─ free-3
├─ mdat-4
Total structure = 258

```

Telegram

```

ftyp-1
├─ @majorBrands: mp42
├─ @minorVersion: 1
├─ @compatibleBrand:
  mp41mp42isom
mdat-2
moov-3
├─ mvhd-1
│   └─ version: 0
│   └─ @flags: 0
│   └─ @reserved: 0
│   └─ @timeScale: 44100
│   └─ @posterTime: 0
│   └─ @currentTime: 0
│   └─ @nextTrackId: 3
│   └─ @previewTime: 0
│   ...
├─ trak-1
│   └─ tkhd-1
│       └─ @version: 0
│       └─ @flags: 1
│       ...
├─ trak-2
Total structure = 251

```

Figura 2: Comparación de la estructura de contenedores multimedia

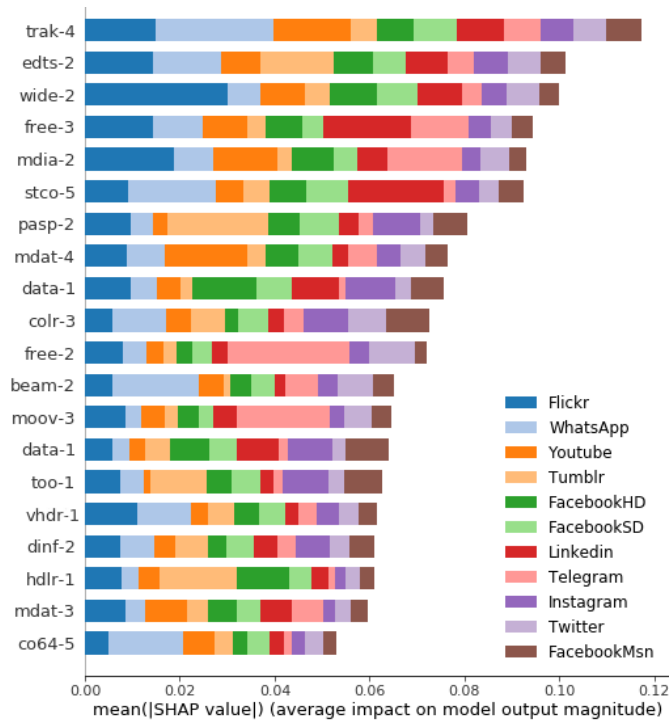


Figura 3: Factor de impacto de *PathOrder-tag*

Los *PathOrder-tag* que contienen el átomo *trak-4* tienen gran impacto para que el modelo clasifique la muestra como WhatsApp. Este experimento nos muestra que el modelo clasifica notablemente las redes sociales. No obstante, recordar que la red social Flickr clasifica perfectamente pero realmente esta clasificando vídeos originales.

En la Figura 4, se examina la predicción individual de los puntos de datos (*PathOrder-tag*) por redes sociales y aplicaciones de mensajería instantánea utilizando el paquete de Python Eli5 [34]. En la Figura 4(a) se examina Facebook Msn, Instagram y Flickr. Se observa que el *PathOrder-tag* /moov-2/udta-4/meta-1/ilst-2/too-1/data-1/ tiene mayor peso de contribución y es el más influyente para predecir la aplicación de mensajería Facebook Msn (+0.056) y la red social Instagram (+0.050). No obstante, el conjunto de *PathOrder-tag* positivos de Facebook Msn, alcanza una probabilidad del 0.753 % para que el resultado de la predicción sea ésta aplicación, mientras que el conjunto de *PathOrder-tag* positivos de Instagram, alcanza una probabilidad de 0.244 % para que el resultado de predicción se incline a ésta red social. Asimismo, se observa que el conjunto de *PathOrder-tag* positivos de Flickr logra obtener solo un 0.002 % de probabilidad para que la predicción sea ésta red social (vídeo original).

y=FacebookMsn (probability 0.753) top features			y=Instagram (probability 0.244) top features			y=Flickr (probability 0.002) top features		
Contribution [?]	Feature	Value	Contribution [?]	Feature	Value	Contribution [?]	Feature	Value
+0.074	<BIAS>	1.000	+0.084	<BIAS>	1.000	+0.138	<BIAS>	1.000
+0.056	/moov-2/udta-4/meta-1/ilst-2/too-1/data-1/	-0.854	+0.050	/moov-2/udta-4/meta-1/ilst-2/too-1/data-1/	-0.854	...	6 more positive ...	
						...	33 more negative ...	
+0.034	/moov-2/trak-2/mdia-3/minf-3/dinf-2/dref-1/	-3.103	+0.028	/moov-2/trak-2/mdia-3/minf-3/dinf-2/dref-1/	-3.103	-0.009	/moov-2/meta-3/ilst-3/	0.002
+0.029	/moov-2/trak-3/mdia-2/mdhd-1/	-0.626				-0.020	/moov-2/udta-4/meta-1/ilst-2/too-1/data-1/	-0.854
+0.027	/moov-2/trak-2/mdia-3/minf-3/stbl-3/stds-1/avc1-1/pasp-2/	-0.145	+0.020	/moov-2/trak-3/mdia-2/mdhd-1/	-0.626	-0.022	/moov-2/trak-3/mdia-2/mdhd-1/	-0.626
	... 39 more positive ...		+0.016	/moov-2/udta-4/meta-1/ilst-2/too-3/data-1/	-0.058	-0.041	/moov-2/trak-2/mdia-3/minf-3/dinf-2/dref-1/	-3.103
				... 20 more positive ...				
				... 19 more negative ...				

(a) Facebook Msn, Instagram, Flickr

y=Instagram (probability 0.998) top features			y=Flickr (probability 0.001) top features			y=FacebookMsn (probability 0.001) top features		
Contribution [?]	Feature	Value	Contribution [?]	Feature	Value	Contribution [?]	Feature	Value
+0.084	<BIAS>	1.000	+0.138	<BIAS>	1.000	+0.074	<BIAS>	1.000
+0.065	/moov-2/udta-4/meta-1/ilst-2/too-1/data-1/	-0.448	...	7 more positive ...		+0.039	/moov-2/udta-4/meta-1/ilst-2/too-1/data-1/	-0.448
			...	32 more negative ...				
+0.042	/moov-2/trak-2/mdia-3/minf-3/stbl-3/stds-1/avc1-1/pasp-2/	-0.497	-0.009	/moov-2/meta-3/ilst-3/	-0.003	+0.021	/moov-2/trak-2/mdia-3/minf-3/dinf-2/dref-1/	-3.089
+0.040	/moov-2/trak-2/mdia-3/minf-3/dinf-2/dref-1/	-3.089	-0.019	/moov-2/udta-4/meta-1/ilst-2/too-1/data-1/	-0.448	+0.013	/moov-2/trak-3/mdia-2/mdhd-1/	-0.610
+0.034	/moov-2/trak-3/mdia-2/mdhd-1/	-0.610	-0.021	/moov-2/trak-3/mdia-2/mdhd-1/	-0.610	...	11 more positive ...	
	... 39 more positive ...		-0.040	/moov-2/trak-2/mdia-3/minf-3/dinf-2/dref-1/	-3.089	...	28 more negative ...	
						-0.012	/moov-4/trak-2/mdia-4/minf-3/stbl-4/stds-1/hvc1-1/	0.000

(b) Instagram, Flickr, Facebook Msn

y=WhatsApp (probability 0.999) top features			y=Flickr (probability 0.001) top features			y=YouTube (probability 0.000) top features		
Contribution [?]	Feature	Value	Contribution [?]	Feature	Value	Contribution [?]	Feature	Value
+0.234	/beam-2/	2.923	+0.138	<BIAS>	1.000	+0.084	<BIAS>	1.000
+0.137	<BIAS>	1.000	+0.021	/moov-2/trak-2/mdia-3/minf-3/dinf-2/dref-1/	3.651	...	19 more positive ...	
+0.105	/moov-2/trak-2/mdia-3/minf-3/dinf-2/dref-1/	3.651	...	8 more positive ...		...	18 more negative ...	
+0.066	/moov-3/trak-4/mdia-2/hdlr-2/	-0.157	...	31 more negative ...		-0.013	/moov-2/trak-3/mdia-2/mdhd-1/	-1.346
+0.047	/moov-2/trak-3/mdia-2/mdhd-1/	-1.346	-0.018	/moov-2/trak-3/mdia-4/minf-3/stbl-4/sdtp-4/	-0.000	-0.019	/moov-2/trak-3/mdia-2/hdlr-2/	-0.011
	... 39 more positive ...		-0.035	/moov-2/trak-3/mdia-2/mdhd-1/	-1.346	-0.023	/beam-2/	2.923
			-0.050	/beam-2/	2.923	-0.029	/moov-2/trak-2/mdia-3/minf-3/dinf-2/dref-1/	3.651

(c) WhatsApp, Flickr, Youtube

Figura 4: Predicción individual de puntos de datos en Redes sociales y aplicaciones de mensajería.

La interpretación de las demás figuras se realiza de forma similar a la anterior. En resumen, una o más clases pueden tener el mismo *PathOrder-tag* con un cierto grado de influencia, pero también es importante la probabilidad que alcanza el conjunto de *PathOrder-tag* positivos, ya que de ambos factores depende una clasificación eficiente de los vídeos con su respectiva clase (Redes sociales y aplicaciones de mensajería instantánea).

VI. CONCLUSIONES

En este trabajo se presenta un método de detección de post-procesamientos en vídeos compartidos a través de redes sociales y aplicaciones de mensajería instantánea que analiza los cambios presentes en la estructura de contenedores multimedia. Para la implementación del método propuesto, en primer lugar, se generó un conjunto de datos de vídeos y posteriormente fueron compartidos a través de las redes sociales y aplicaciones de mensajería instantánea más usados en la actualidad. En segundo lugar, se realizó la extracción del conjunto de características (*PathOrder-tag*) del 100 % de vídeos con el algoritmo de extracción de átomos. En tercer lugar, se construyó un modelo de machine learning supervisado a partir de la evaluación de 4 métricas obtenidas por 9 algoritmos de clasificación, siendo *Random Forest (RF)* el algoritmo que obtuvo el mejor rendimiento. En el análisis de los contenedores multimedia se demostró que cada red social y aplicación de mensajería inserta una estructura propia a los contenedores multimedia, excepto Flickr, ya que no realiza un proceso de re-compresión sobre vídeo al momento de ser compartido por su plataforma. Las diferencias detectadas en cuanto a número de átomos es considerable (Facebook HD: 267, Youtube: 250, WhatsApp: 237, LinkedIn: 258, Telegram: 251), y en cuanto a valores es mucho más evidente.

En el experimento de identificación de redes sociales y aplicaciones de mensajería instantánea se alcanzó el 100 % de acierto para todas las clases, excepto Instagram que se confunde un 0.31 % con Facebook Msn y Facebook SD se confunde un 0.02 % con Tumblr.

En líneas generales, el método propuesto utilizando la estructura de contenedores multimedia es eficiente y robusto para detectar las redes sociales y aplicaciones de mensajería instantánea en vídeos con formato MP4 y MOV.

AGRADECIMIENTOS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700326. Website: <http://ramses2020.eu>. This paper has also received funding from THEIA (Techniques for Integrity and authentication of multimedia files of mobile devices) UCM project (FEI-EU-19-04).



REFERENCIAS

- [1] CISCO, "Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper," February 2019.
- [2] HOOTSUITE, "Global Digital Reports 2019," pp. 1–221, February 2019.
- [3] J. Kaur and N. Sharma, "Survey on the General Concepts of MPEG Moving Picture Experts Group," *Paripex: Indian Journal of Research*, vol. 5, no. 2, pp. 252–255, February 2016.
- [4] B. G. Haskell, P. A., and N. A. N., *Digital Video: An Introduction to MPEG-2 Digital Multimedia Standards*. Orlando, FL, USA: Springer US, 2007.
- [5] S. Dhanani and M. Parker, *Digital Video Processing for Engineers: A Foundation for Embedded Systems Design*. Newton, MA, USA: Newnes, 2012.
- [6] I. T. Union, "Advanced Video Coding for Generic Audiovisual Services H.264," 2016. [Online]. Available: <http://www.itu.int/>
- [7] International Telecommunication Union, "High Efficiency Video Coding," 2018. [Online]. Available: <http://www.itu.int/>
- [8] Google, "Android-OS," 2019. [Online]. Available: <https://www.android.com/>
- [9] Q. F. Format, "QuickTime File Format Specification," 2016. [Online]. Available: <https://developer.apple.com>
- [10] Apple Inc., "iOS," 2019. [Online]. Available: www.apple.com/es/ios
- [11] Microsoft, "AVI RIFF File," 1992. [Online]. Available: <https://docs.microsoft.com/en-us/windows/win32/directshow/avi-riff-file-reference>
- [12] Matroska, "Matroska Specification," 2002. [Online]. Available: <https://www.matroska.org/technical/specs/index.html>
- [13] Tengku Mohd T. S. and Halimah Badioze Z. and Hsinchun C. and Shalini R. U. and Sung Hyon M., "Digital libraries: Technology and management of indigenous knowledge for global access," in *Proceedings of the 6th International Conference on Asian Digital Libraries*. Springer, Berlin, Heidelberg, December 2003, pp. 76–83.
- [14] F. Camastra and A. Vinciarelli, *Machine Learning for Audio, Image and Video Analysis*.
- [15] J. Lukas, J. Fridrich, and M. Goljan, "Digital Camera Identification from Sensor Pattern Noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, June 2006.
- [16] C. Li, "Source Camera Identification Using Enhanced Sensor Pattern Noise," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, June 2010.
- [17] Z. J. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, and N. Saitoh, "Methods for Identification of Images Acquired with Digital Cameras," in *Proceedings on Enabling Technologies for Law Enforcement and Security*, Boston, Massachusetts, USA, February 2001.
- [18] Van Lanh, T. and Chong, K. S. and Emmanuel, S. and Kankanhalli, M. S., "A Survey on Digital Camera Image Forensic Methods," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, Beijing, July 2007, pp. 16–19.
- [19] Costa, F. O. and Eckmann, M. and Scheirer, W. J. and Rocha, A., "Open Set Source Camera Attribution," in *Proceedings of the 25th Conference on Graphics, Patterns and Images*, Ouro Preto, Brazil, August 2012, pp. 71–78.
- [20] Li, J. and Ma, B. and Wang, C., "Extraction of PRNU Noise From Partly Decoded Video," *Journal of Visual Communication and Image Representation*, vol. 57, pp. 183–191, November 2018.
- [21] I. Amerini, R. Caldelli, A. Del Castillo, A. Di Fuccia, C. Molinari, and A. P. Rizzo, "Dealing with Video Source Identification in Social Networks," *Signal Processing: Image Communication*, vol. 57, pp. 1–7, September 2017.
- [22] A. Sandoval Orozco, D. Arenas González, J. Rosales Corripio, L. García Villalba, and J. C. Hernandez-Castro, "Techniques for Source Camera Identification," in *Proceedings of the 6th International Conference on Information Technology*, Amman, Jordan, May 2013, pp. 1–9.
- [23] J. Song, K. Lee, W. Y. Lee, and L. H., "Integrity Verification of the Ordered Data Structures in Manipulated Video Content," *Digital Investigation*, vol. 18, no. C, pp. 1–7, Septiembre 2016.
- [24] Gloe, T. and Fisher, A. and Kirchner, M., "Forensic Analysis of Video File Formats," in *Proceedings of the First Annual DFRWS Europe*, Munster, Germany, May 2014, pp. 68–76.
- [25] M. Iuliani, D. Shullani, M. Fontani, M. S., and A. Piva, "A video forensic framework for the unsupervised analysis of mp4-like file container," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 3, pp. 635–645, March 2018.
- [26] Annies, S., "MP4 Parser." [Online]. Available: <https://github.com/sannies/mp4parser>
- [27] S. Maldonado, J. Lopez, and C. Vairetti, "An alternative smote oversampling strategy for high-dimensional datasets," *Applied Soft Computing*, vol. 76, pp. 380–389, 2019.
- [28] D. Freeman and C. Chio, *Machine Learning and Security*. Boston, USA: O'Reilly Media, February 2018.
- [29] KAGGLE, "KAGGLE: Online Community of Data Scientists and Machine Learners." [Online]. Available: <https://www.kaggle.com/>
- [30] G. Rebala, A. Ravi, and S. Churiwala, *An Introduction to Machine Learning*, 1st ed. Switzerland: Springer, January 2019.

- [31] D. Sarkar, R. Bali, and T. Sharma, *Practical Machine Learning with Python*, 2nd ed. Massachusetts, USA: Apress, December 2017.
- [32] S. Rachka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*, 2nd ed. Birmingham, United Kingdom: Packt Publishing, September 2017.
- [33] S. Guido and A. C. Muller, *Introduction to Machine Learning with Python*. USA: O'Reilly Media, December 2017.
- [34] ELI5, "ELI5 for Python," 2019. [Online]. Available: <https://eli5.readthedocs.io/en/latest/overview.html>

Carlos Quinto Huamán received his Computer Science degree in 2012 at Universidad Inca Garcilaso de la Vega in Lima (Perú) and a M.Sc. degree in Computer Science in 2016 from the Universidad Complutense de Madrid (Spain). He is currently a Ph.D. student in the Department of Software Engineering and Artificial Intelligence of the Faculty of Computer Science and Engineering at the Universidad Complutense de Madrid (UCM) and member of the Complutense Research Group GASS (Group of Analysis, Security and Systems, <http://gass.ucm.es>). His research interests are: computer forensics, cybersecurity, electronic warfare and cyberdefense.

Daniel Povedano Álvarez received his Computer Science Engineering degree in 2017 at Universidad Complutense de Madrid. He is currently a M.Sc. student of Data Science in Universidad Complutense de Madrid and member of the Complutense Research Group GASS (Group of Analysis, Security and Systems, <http://gass.ucm.es>). His research interests are: computer forensics, data science, cybersecurity, artificial intelligence.

Ana Lucila Sandoval Orozco was born in Chivolo, Magdalena, Colombia in 1976. She received a Computer Science Engineering degree from the Universidad Autónoma del Caribe (Colombia) in 2001. She holds a Specialization Course in Computer Networks (2006) from the Universidad del Norte (Colombia), and holds a M.Sc. in Research in Computer Science (2009) and a Ph.D. in Computer Science (2014), both from the Universidad Complutense de Madrid (Spain). She is currently a postdoctoral researcher and member of the Research Group GASS (Group of Analysis, Security and Systems, <http://gass.ucm.es>) at Universidad Complutense de Madrid (Spain). Her main research interests are coding theory, information security and its applications.

Luis Javier García Villalba received a Telecommunication Engineering degree from the Universidad de Málaga (Spain) in 1993 and holds a Ph.D. in Computer Science (1999) from the Universidad Politécnica de Madrid (Spain). Visiting Scholar at COSIC (Computer Security and Industrial Cryptography, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, Belgium) in 2000 and Visiting Scientist at IBM Research Division (IBM Almaden Research Center, San Jose, CA, USA) in 2001 and 2002, he is currently Associate Professor of the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM) and Head of Complutense Research Group GASS (Group of Analysis, Security and Systems) which is located in the Faculty of Computer Science and Engineering at the UCM Campus. His professional experience includes the management of both national and international research projects and both public (Spanish Ministry of R&D, Spanish Ministry of Defence, Horizon 2020 - European Commission, . . .) and private financing (Hitachi, IBM, Nokia, Safelayer Secure Communications, TB Solutions Security, . . .). Author or co-author of numerous international publications is editor or guest editor of numerous journals such as Entropy MPDI, Future Generation Computer Systems (FGCS), Future Internet MDPI, IEEE Latin America Transactions, IET Communications (IET-COM), IET Networks (IET-NET), IET Wireless Sensor Systems (IET-WSS), International Journal of Ad Hoc and Ubiquitous Computing (IJAHUC), International Journal of Multimedia and Ubiquitous Engineering (IJMUE), Journal of Supercomputing, Sensors MDPI, etc.

Técnica de Detección de Manipulación en Vídeos Digitales Basada en los Algoritmos de Compresión

Edgar González Fernández, Ana Lucila Sandoval Orozco, and Luis Javier García Villalba, *Member, IEEE*

Resumen—Las imágenes y vídeos digitales juegan un papel muy importante en la vida cotidiana. A día de hoy, la mayor parte de la población es poseedora de cámaras fotográficas de última generación integradas en su dispositivo móvil. El desarrollo tecnológico no sólo facilita la generación de contenido multimedia, sino también la manipulación intencionada de éste, y es aquí donde las técnicas forenses de detección de manipulación sobre imágenes y vídeos cobran gran importancia. En este trabajo se proponen dos metodologías forenses basadas en algoritmos de compresión: La primera de ellas trata de detectar la presencia de recompresión en un vídeo digital mediante el análisis de sus macrobloques, característica propia del estándar H.264-MPEG4. Posteriormente, se utiliza la máquina de soporte vectorial para crear el modelo que permita la verificación del número de recompresiones de un vídeo. La segunda metodología que se explica en este trabajo tiene por objetivo detectar alteraciones de tipo ‘empalme’, es decir, regiones que no pertenecen al contenido original de una imagen digital, técnica que está basada en la tasa de error que introduce el algoritmo de compresión JPEG cada vez que recomprime una imagen.

Palabras claves—Análisis Forense, Clasificación, Compresión, Macrobloques, Manipulación, Máquinas de Vector Soporte, Vídeos Digitales.

I. INTRODUCCIÓN

Desde siglos atrás, el ser humano siempre ha utilizado la imagen para plasmar la realidad que le rodeaba, o modificarla, en función del mensaje que se quisiera transmitir. Aunque esta evolución, sin duda, tiene un antes y un después con la creación de la fotografía en el siglo XIX.

“La excitación que acompañó a la invención de la fotografía fue la sensación de que el hombre por primera vez podía ver el mundo como realmente era”(Collier 1986: 3) [1].

Esta afirmación que hace Collier acerca de la fotografía podría no ajustarse al pie de la letra en la actual era digital. Actualmente existe un significativo número de delitos informáticos relacionados con la posesión ilícita, distribución o modificación de contenido multimedia. El uso de dispositivos móviles para este propósito hace de estos una importante fuente de evidencia, hecho por el cual los análisis forenses deben ser capaces de autenticar el contenido y examinar si es original o fue manipulado.

La facilidad para manipular imágenes y vídeos digitales se ha incrementado vertiginosamente en los últimos tiempos, y está al alcance del usuario convencional mediante programas como Adobe Photoshop, GIMP, Adobe Premiere, etc. Incluso, estas manipulaciones son realizadas de manera automática

en dispositivos móviles mediante nuevas herramientas que hacen uso de la inteligencia artificial, como pueden ser los embellecedores de rostros, cambios de la expresión facial, mejora de la iluminación de la escena, etc.

En Julio del año 2017 los investigadores de la revista Cognitive-Research [2] utilizaron un dataset de 40 escenas, 30 de las cuales fueron sometidas a cinco tipos diferentes de manipulación, incluyendo manipulaciones físicamente plausibles y no plausibles. Se mostraron a 707 participantes con el fin de evaluar la capacidad de las personas para detectar escenas manipuladas del mundo real. El estudio encontró que sólo el 60 % de las personas fue capaz de detectar las escenas falsas, e incluso entonces, sólo un 45 % de ellos fueron capaces de decir dónde exactamente se encontraba la alteración del contenido (ver Figura 1).

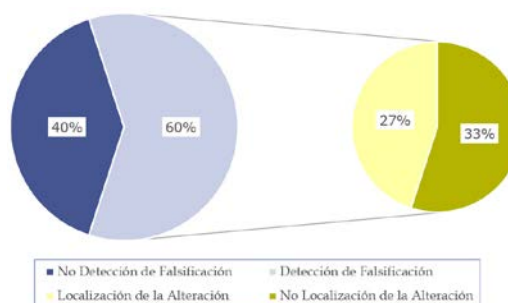


Figura 1: Resultados del estudio de la revista Cognitive-Research.

Es por estos motivos que deben desarrollarse técnicas de detección de manipulación en imágenes y vídeos, imprescindibles para dicho fin. Se hace necesaria la revisión y mejora de los métodos de verificación de la autenticidad e integridad del contenido de una imagen o vídeo, así como el desarrollo de nuevos métodos dirigidos a las técnicas que en el futuro puedan plantearse.

El presente trabajo se desarrolla como sigue: En la Sección II se da una breve introducción a las técnicas de manipulación en vídeos digitales. En la Sección III se presenta el estado del arte sobre detección de manipulaciones en vídeos. Los algoritmos propuestos así como los conceptos necesarios para comprenderlos son explicados en la Sección IV. Posteriormente, en la Sección V se muestran los experimentos realizados y los resultados obtenidos. Finalmente, en la Sección VI se recogen las conclusiones y se proponen trabajos futuros en este campo respectivamente.

E. González Fernández, A. L. Sandoval Orozco and L. J. García Villalba son miembros del Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Ingeniería del Software e Inteligencia Artificial (DISIA), Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid e-mail: edggonza@ucm.es, {asandoval, javiergv}@fdi.ucm.es.

II. TÉCNICAS DE MANIPULACIÓN EN VÍDEOS

II-A. Inter-Fotograma

Un vídeo digital se compone de una secuencia de imágenes llamadas fotogramas. Las manipulaciones de tipo inter-fotograma (inter-frame) se centran en la modificación de la correlación temporal entre ellos. Para modificar la correlación temporal del vídeo es posible insertar, duplicar, intercambiar o eliminar cualquiera de los fotogramas que lo conforman (Figura 2).

Otra forma de manipular un vídeo inter-fotograma es mediante el empalme de dos o más vídeos, es decir, interpolando fotogramas de ambos para generar uno nuevo. Además, es posible que los vídeos originales no compartan los mismos fotogramas por segundo (fps), por lo que será necesario también manipular esta característica para ajustar los fps de uno al otro.

El principal objetivo de esta manipulación es el de eliminar de la escena grabada un evento indeseado. También es posible incriminar en la escena a otros objetos con la adición de un fotograma externo. Si se toma como ejemplo la secuencia de las imágenes de vigilancia de una cámara de tráfico, como las de la Figura 3, es sencillo hacer que el vehículo blanco de la Figura 3d desaparezca de la escena eliminando ese fotograma.

En general, el ojo humano no puede detectar diferencias entre el vídeo original y el vídeo con manipulación inter-fotograma pero las operaciones de procesamiento de la manipulación dejan una huella en la información del contenido.

II-B. Intra-Fotograma

La manipulación intra-fotograma se centra en la alteración de cada fotograma individualmente. Estas manipulaciones pueden clasificarse en:

- **Manipulación a nivel de píxel:** La cual consiste en tratar al fotograma como una imagen individual y aplicar técnicas de manipulación en imágenes como las vistas en la sección anterior, por ejemplo, copia-pegar o empalmes.
- **Manipulación a nivel de fotograma:** Mediante la cual se cambia de tamaño o se recortan las extremidades de un fotograma con el objetivo de ocultar cierto contenido del vídeo que se ubique en los bordes del fotograma. Por ejemplo marcas de Hora y lugar de grabación.

A diferencia del ejemplo expuesto con las técnicas inter-fotograma, si se tiene como objetivo ocultar el paso de un vehículo de la cámara de vigilancia de la Figura 3 con técnicas intra-fotograma, en lugar de eliminar el fotograma en el que aparece, se podría hacer desaparecer con técnicas de copia-pegar o incluso se podría re-escalar el fotograma y recortar la zona en la que aparece.

III. TRABAJOS PREVIOS

La Figura 4 Muestra una clasificación de las técnicas de detección de manipulaciones en vídeos digitales.

III-A. Detección de Manipulación Inter-Fotograma

Los dispositivos introducen un ruido en cada fotograma cuando graban un vídeo. Dado que este ruido sigue un patrón particular en una secuencia de fotogramas consecutivos, es

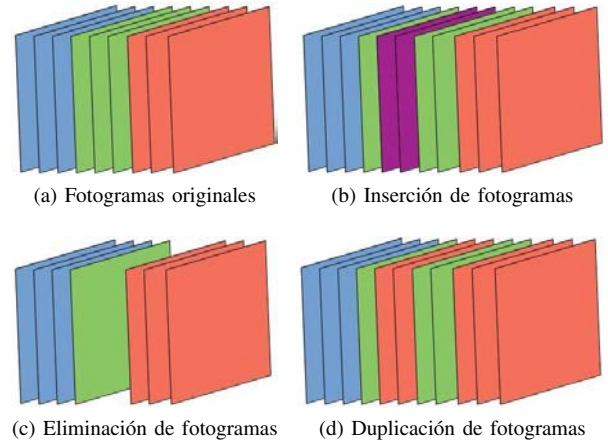


Figura 2: Ejemplo de Manipulación Inter-Fotograma.

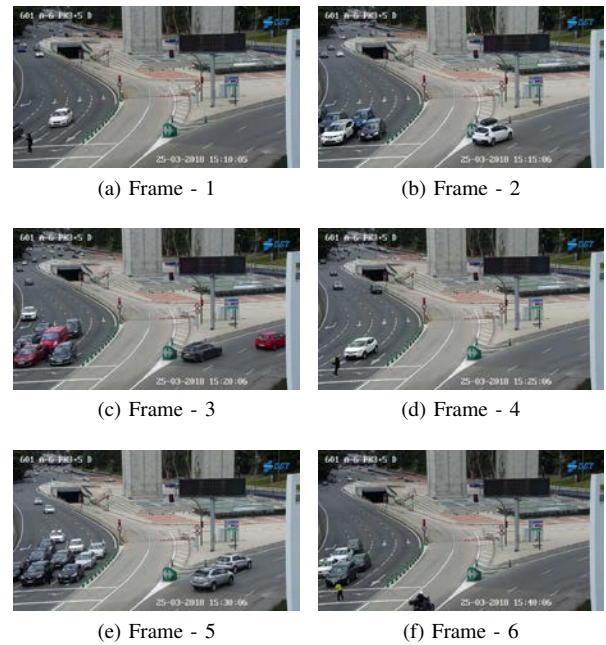


Figura 3: Ejemplo de fotogramas de una cámara de vigilancia de la Dirección General de Tráfico (DGT).

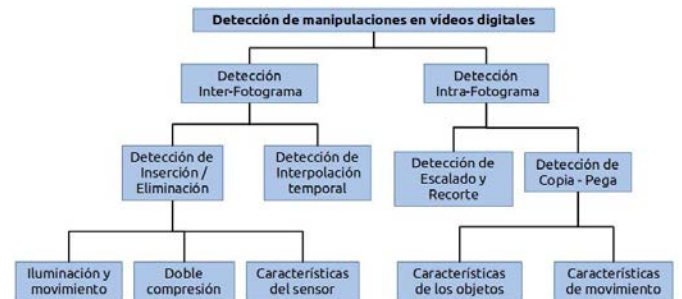


Figura 4: Esquema de Detecciones de Manipulación en Vídeos.

posible que, investigando estos rastros, se detecten cambios entre los fotogramas.

En [3] los autores utilizaron la varianza entre el ruido promedio de los fotogramas y uno en particular. Los fotogramas

con varianzas más altas serían marcados como inserciones. No se demostró su eficacia sobre vídeos comprimidos, y además, su eficacia fue probada sobre vídeos auto-grabados y no es suficiente para determinar su aplicabilidad.

Los autores de [4] propusieron un algoritmo de movimiento adaptativo que fue capaz de detectar y localizar falsificaciones en vídeos entrelazados y descentrelazados. Se basaban en la detección de las perturbaciones en la correlación para los entrelazados y en los disturbios de movimiento entre fotogramas para los descentrelazados. Este método, no obstante, resultaba ineficaz para vídeos de baja calidad.

En [5] utilizan el concepto del Patrón de Ruido del Sensor (SPN) de la cámara para determinar si todos los fotogramas del vídeo habían sido grabados con el mismo dispositivo. Los resultados obtenidos indicaron que el algoritmo era fiable para vídeos no comprimidos, pero el rendimiento se deterioraba para vídeos comprimidos.

Otra propuesta basada en la cámara es presentada en [6], que proporcionó autenticación a nivel de píxel para todos los fotogramas del vídeo. Se basaba en las inconsistencias en los fotones del ruido de disparo que introduce la cámara durante el proceso de adquisición. También lograrían encontrar regiones sospechosas en los fotogramas.

Otra forma de manipular un vídeo es mediante un corte temporal, intercalando fotogramas de dos vídeos diferentes. Cuando se intercalan dos fotogramas hay que tener en cuenta que es necesario sincronizar sus velocidades (frame-rate).

El método sugerido en [7] se basa en la propiedad de interpolación compensada por movimiento ya que deja huellas detectables en los fotogramas. Los autores pudieron sugerir un sistema que funcionaba para vídeos no comprimidos y ligeramente comprimidos (por ejemplo, H.264, o vídeos de transmisión de televisión) y lograba resultados prometedores, incluso cuando se usaba sólo en un subconjunto de fotogramas. Además, el sistema funcionaba bien en ventanas espaciales de pequeño tamaño, lo que permitió que este detector se usara como una posible herramienta para detectar ataques de falsificación de copiar y pegar. Sin embargo, el número de cuadros interpolados observados tenía que ser lo suficientemente grande para que el sistema detectara las falsificaciones con éxito.

En [8] se detecta la conversión de velocidad ascendente de fotogramas basándose en la intensidad de los bordes. Utilizan un umbral determinado para distinguir las zonas originales de las convertidas al alza, y en base a ello, estiman la velocidad teórica de los fotogramas originales. Para un total de 300 secuencias de prueba, consiguieron un promedio de detección de 95 %.

Los autores en [9] desarrollaron un método de detección ciego basado en el análisis a nivel de fotograma de una característica llamada “variación media de la textura” (ATV). Cada curva ATV generada se procesaba en el vídeo candidato como evidencia de la conversión de velocidad ascendente. Esta técnica podría localizar la posición de la interpolación de los fotogramas y ayudar a estimar su velocidad original.

III-B. Detección de Manipulación Intra-Fotograma

Las técnicas de detección de copia-pegar proceden buscando similitudes entre regiones de fotogramas sucesivos o dentro del mismo fotograma.

Para detectar estas falsificaciones, los autores en [4] calcularon coeficientes de correlación espacial y temporal para identificar y localizar semejanza entre partes separadas del vídeo. Este método, obtuvo muy buenos resultados para vídeos con compresión MPEG debido a que los artefactos de compresión son más pronunciados en presencia de movimiento en el vídeo.

Otra técnica propuesta en [10] y [11] se basó en la hipótesis de que los atributos de correlación de sub-bloques de píxeles intra e inter fotograma están obligados a ser desorganizados debido a alteraciones como doble compresión, retoque o remuestreo. Los autores extrajeron los residuos de ruido y la cuantificación de características de fotogramas adyacentes para luego realizar un análisis de correlación usando el análisis de correlación canónico, análisis factorial intermodal, y análisis semántico latente. Tales perturbaciones ayudaron a la técnica para diferenciar las huellas de un vídeo original de las de uno manipulado.

En [12], se propone detectar alteraciones con la conversión del vídeo a una secuencia de fotogramas, seguido de un proceso de emparejamiento de bloques dentro de la región sospechosa. Al trabajar en una parte del fotograma en lugar de todo el fotograma, la técnica es capaz de mantener un buen equilibrio entre rendimiento y complejidad.

El método de detección y localización de falsificación de [13] era similar en funcionalidad a [14] pero de manera completamente automática. Es un algoritmo de dos pasos, en el que primero se detectan manipulaciones a nivel de fotograma y se analiza el vídeo residual que se obtiene al restar píxeles que ocupan la misma posición espacial en fotogramas consecutivos. Entonces, para detectar el contenido duplicado, los autores ponen en relación los bloques 3D de los fotogramas. Así, la presencia de alta correlación indica la ubicación del contenido idéntico.

En [15], los autores propusieron un enfoque para detectar y localizar falsificaciones a nivel de región en vídeos. El método detecta irregularidades en la coherencia espacio-temporal entre fotogramas consecutivos. El vídeo primero se divide en conjuntos de fotogramas y luego se calcula la coherencia entre cada una de estos conjuntos. Dicho así, los conjuntos con coherencia antinaturalmente alta o coherencia anormalmente baja se clasificarían como fotogramas manipulados.

Una técnica de localización y detección de eliminación de objetos es la que se nos presenta en [16]. Se utilizó aquí Scale-Invariant Feature Transform (SIFT) junto con la coincidencia k-NN y correlación cruzada ruido-residuo para detectar falsificaciones copia-pegar. Aunque la técnica funciona bien para la prueba vídeos, esta sufre una degradación significativa a medida que aumenta la resistencia a la compresión.

Otra forma de manipular el contenido del vídeo es ampliando un fotograma y después eliminando el evento incriminatorio recortando la parte externa de éste. Es importante saber que cuando se hace un recorte se produce un remuestreo para mantener una resolución constante en todos los fotogramas del vídeo.

En [17] los autores observaron que el remuestreo introduce ciertas correlaciones estadísticas sobre el contenido dado. Explotaron el SPN como característica forense y analizaron las variaciones en las propiedades de correlación de referencia SPN y el de re-escalado. Este método es bastante firme en

cualquier tipo de vídeo, pero también resulta excesivamente dependiente de una gran cantidad de parámetros y umbrales dependientes del contenido, lo cual requiere un ajuste empírico extremadamente cuidadoso.

III-C. Detección de Doble Compresión

La recompresión o doble compresión es, una consecuencia inevitable de la falsificación, y su detección podría ayudar a detectar la presencia manipulaciones.

Los primeros pasos en esta dirección se pueden atribuir a los autores de [18]. Su algoritmo se basaba en la suposición simple de que cuando se manipulaba un vídeo MPEG, se producían dos compresiones: primero, cuando se creaba el vídeo y, segundo, cuando se volvían a guardar después de dicha alteración. También explotaron el hecho de que dentro de un Grupo de Imágenes (GOP), los fotogramas muestran una gran correlación entre ellos, de manera que al agregar o eliminar un fotograma en un GOP aumenta el error de estimación de movimiento, lo que también da como resultado picos periódicos detectables.

En [19], los autores presentaron una técnica para detectar la cuantización doble, que resultó de la recompresión de un vídeo comprimido Moving Picture Experts Group (MPEG) o de la combinación de vídeos de características diferentes. La técnica podría detectar una manipulación si los coeficientes Transformada Discreta del Coseno (DCT) de los fotogramas del vídeo se sometieron a doble compresión en cualquier punto. Los resultados empíricos indicaron que la tasa de detección fue altamente dependiente de la relación de la primera y la segunda escala de cuantificación. Evidentemente, la técnica fue efectiva siempre que el segundo factor de calidad de compresión fuera más alto que el primero.

En [20], las falsificaciones en vídeos codificados en MPEG-2 se detectaron mediante el examen de la distribución del coeficiente DCT. Este algoritmo se basó en la observación de que el histograma de coeficientes de DCT cuantificados de un vídeo que había experimentado una doble compresión exhibía un patrón convexo. A diferencia de [19], que dependía en gran medida de las escalas de cuantificación, los autores en este caso sugirieron controlar la tasa de bits de salida, lo que hizo que este algoritmo se adaptara a las necesidades de diferentes tipos de sistemas de codificación de vídeo pero no pudo localizar la falsificación en el vídeo. Tampoco pudo funcionar bien para vídeos de cámara lenta.

El trabajo en [21] también se centró en la detección de alteraciones basadas en fotogramas al detectar la compresión doble en vídeos MPEG-2. En lugar de basar el proceso de detección de agregación/eliminación de trama en las características temporales, los autores sugirieron utilizar las características de frecuencia. Se observó que cuando se vuelve a comprimir un vídeo MPEG después de agregar/eliminar el fotograma, se pierden algunos componentes de alta frecuencia en los fotogramas recomprimidos debido a la descincronización de los GOP y la cuantificación no lineal realizada en el proceso de codificación. Estas variaciones no solo ayudan a detectar la falsificación sino también a localizarla.

Otra técnica de detección de falsificación basada en doble compresión MPEG es la propuesta en [22], donde las anomalías en los patrones de coeficientes DCT son tratadas

como indicativo de inserción/eliminación de fotogramas. Los autores extrajeron características de los GOP, que luego son utilizados por una Máquina de Soporte Vectorial (SVM) para determinar la velocidad de bits original del vídeo doblemente comprimido dado, y se observa que el rendimiento de detección de la técnica era relativamente inferior para los vídeos con menor tasa de bits, porque una escala de cuantificación mayor requiere un proceso de cuantificación más robusto, que la técnica no estaba preparada para manejar.

En el mismo año, se pre-planteó una técnica similar en [23] aunque con una novedad: su capacidad para detectar vídeos transcodificados, es decir, vídeos que habían sido doblemente comprimidos utilizando dos estándares de compresión diferentes. Los autores observaron además que después de que un vídeo MPEG-2 se transformara en vídeo MPEG-4, las trazas de compresión MPEG-2 anteriores, estos generan nuevas periodicidades que se observaron claramente en los histogramas de los coeficientes DCT reconstruidos. Los autores presentaron los resultados en forma de curvas de características operativas del receptor y declararon que se habían obtenido resultados perfectos en caso de bajas tasas de bits. Estas curvas también demostraron que a medida que aumentaba la velocidad de bits de salida objetivo, el rendimiento de detección disminuía. También asumió que la transcodificación siempre sugería manipulación.

Por otra parte, en [24] se propone detectar la codificación doble incluso si el conjunto de fotogramas principales hubiera sido eliminado. Este método tiene la ventaja adicional de poder ubicar efectivamente la falsificación, además de resultar adecuado también para vídeos codificados H.264, a diferencia de [23] que funcionaba solo para vídeos MPEG. La metodología modificada también fue capaz de estimar el número de fotogramas borrados.

Los autores en [25] declararon que las compresiones múltiples eran un tema poco explorado y que era arriesgado hacer suposiciones con respecto a la autenticidad del contenido digital simplemente sobre la base de la presencia de doble compresión. Su afirmación fue respaldada por el simple hecho de que el contenido digital disponible en Internet, generalmente, sufre más de una compresión.

En [26] utilizan las estadísticas de Markov para detectar doble compresión. Se basan en que la cuantización doble con diferentes parámetros inevitablemente introducirá errores de redondeo, dejando artefactos detectables. El proceso aleatorio de Markov podría capturar dichos artefactos para la detección.

En [27] se basan en características estadísticas de los macrobloques de los P-fotogramas. Proponen detectar la compresión doble de MPEG con el mismo QS. La extracción de características se produce durante la compresión repetida del vídeo en el mismo factor de calidad.

En [28] analizan la degradación que se produce durante una recompresión encontrando que la variación de las características de un vídeo tienden a estabilizarse tras múltiples recompresiones.

En [29] estudian los efectos de la recompresión en los fotogramas predictivos para generar un vector de características con el cual detectar doble compresión a nivel GOP.

Para un vídeo ordinario (descargado de Internet o grabado con ciertos dispositivos móviles), la presencia de signos de doble compresión puede no ser sospechosa pero tampoco

debe considerarse inocua. Si se supone que un vídeo ha sido inalterado, la doble compresión no debería aparecer en dicho vídeo. Por otro lado, si un vídeo dado muestra signos de doble compresión, indicaría la presencia de algún tipo de modificación no autorizada.

Por lo tanto, la presencia de signos de doble compresión serviría como primera, y posiblemente, más importante evidencia de alteración en vídeos.

IV. DETECCIÓN DE DOBLE COMPRESIÓN EN VÍDEOS

IV-A. Conceptos Generales

Para una correcta comprensión de la explicación del algoritmo de detección de doble compresión primero es importante conocer las características del formato de vídeo H.264 y de las herramientas FFMPEG y LIBSVM, utilizadas ambas por dicho algoritmo.

IV-A1. El Formato H.264/MPEG4: Este estándar de codificación fue desarrollado con el objetivo de mejorar la calidad de la imagen, mejorar la eficiencia de codificación y mejorar la robustez de errores en comparación con normas anteriores como MPEG-2, H.263, etc...

El diseño de codificación de este estándar está basado en bloques, es decir, cada fotograma codificado se representa como una unidad de bloques llamados macrobloques. El algoritmo de codificación es el conjunto que se forma al predecir fotogramas por medio de esos macrobloques para explotar dependencias estadísticas temporales, y al transformar la predicción residual para explotar las dependencias estadísticas espaciales.

Algunas de las características más destacadas del diseño, y que además permiten una mayor eficacia de codificación, incluyen las siguientes mejoras en la capacidad para predecir los valores del contenido del fotograma que se va a codificar:

- Por una parte, el tamaño de bloque de compensación de movimiento variable: Este estándar admite más flexibilidad en la selección de tamaños y formas de los bloques, con un tamaño de bloque mínimo de 4:4.
- Por otra, una referencia múltiple para la compensación de movimiento de un fotograma: Los fotogramas con codificación predictiva, llamados P-fotogramas, en estándares anteriores usan sólo el fotograma previo para predecir los valores del fotograma entrante. Este modelo extiende la codificación eficiente al permitir que un codificador seleccione, para fines de compensación de movimiento, entre un mayor número de fotogramas que se han decodificado y almacenado en el decodificador.

El ojo humano percibe el contenido de una escena en términos de información de brillo y color por separado, y con mayor sensibilidad a la de brillo que la de color. El formato H.264 separa una representación de color en tres componentes llamados Y, Cb y Cr. El componente “Y” representa el brillo, mientras que los dos componentes de color “Cb” y “Cr” representan la medida en la que el color se desvía del gris hacia azul y rojo, respectivamente. Debido a que el sistema visual humano es más sensible al brillo que al color, H.264 utiliza una estructura de muestreo en la que el componente cromático tiene un cuarto del número de muestras que el componente lumínico.

Cada fotograma se divide en macrobloques de tamaño fijo que cubren un área rectangular de 16:16 muestras de la componente de brillo y 8:8 muestras de cada uno de los dos componentes de color. Los macrobloques son los componentes básicos para el que se especifica el proceso de decodificación. Todas las muestras de luminancia y croma de un macrobloque se predicen espacial y temporalmente. La señal de vídeo de entrada se divide en macrobloques, cuya asociación se realiza en base a los tipos de fotogramas a los que pertenecen, y luego se procesa cada macrobloque de cada tipo. Es posible un procesamiento en paralelo eficiente [30].

En la compresión .H264 se puede seleccionar la predicción de los macrobloques de manera individual, en lugar de ser los mismos para todo el fotograma, como se observa en la Figura 5):

- **Fotogramas - I:** Todos los macrobloques del fotograma son codificados usando intra-predicción, es decir, no utiliza la información codificada de otros fotogramas.
- **Fotogramas - P:** Además de la codificación de intra-predicción. También se pueden codificar usando inter-predicción con como máximo una señal de predicción de compensación de movimiento por bloque de predicción, es decir, su información proviene del fotograma previo.
- **Fotogramas - B:** Además de los tipos de codificación disponibles en un fotograma P, algunos macrobloques del fotograma B también se pueden codificar utilizando la inter-predicción con dos señales de predicción de compensación de movimiento por bloque de predicción, es decir, su información proviene del fotograma previo y del siguiente.

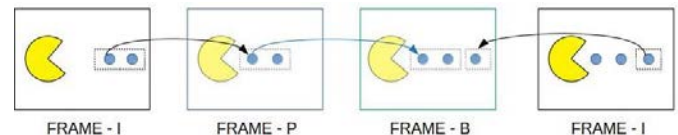


Figura 5: Secuencia de predicción de fotograma.

La señal de predicción para cada macrobloque de $N \times M$ codificado predictivamente se obtiene desplazando un área de la imagen de referencia correspondiente, que se especifica mediante un vector de movimiento de traslación y un índice de referencia de imagen. Los componentes del vector de movimiento se codifican de forma diferencial usando predicción mediana o direccional de bloques vecinos. Ninguna predicción del componente del vector de movimiento (o cualquier otra forma de predicción) tiene lugar a lo largo de los límites del fotograma. En la Figura 6 se muestra un ejemplo de cómo actúa el vector de movimiento en la predicción [30].

IV-A2. La Herramienta FFMPEG: FFMPEG es una plataforma de software libre multimedia capaz de decodificar, codificar, transcodificar, transmitir, filtrar y reproducir la mayoría de formatos de audio y vídeo. Está desarrollado en GNU/Linux pero también compila y ejecuta en la mayoría de sistemas operativos, entornos de desarrollo, arquitecturas y configuraciones. Tiene una licencia GNU LGPL, la cual garantiza una cierta libertad a la hora de compartir y modificar el software, asegurando que el software es libre para todos sus usuarios [31]. Es posible utilizar FFMPEG para analizar los macrobloques y vectores de movimiento de cualquier archivo

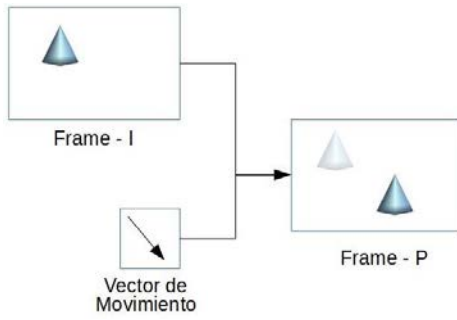


Figura 6: Secuencia de predicción de un fotograma.

de vídeo MP4. En la Figura 7 se puede ver un ejemplo de un fotograma con los vectores de movimiento analizados impresos en forma de flechas [32].



Figura 7: Análisis de los vectores de movimiento.

IV-A3. La Máquina de Soporte Vectorial: Las SVM son técnicas supervisadas de aprendizaje automático, muy útiles para la resolución de problemas de reconocimiento de patrones y para el análisis de regresión. A partir de un conjunto de muestras la SVM construye un modelo que se utiliza para predecir la clase a la que pertenece una nueva muestra. El objetivo de la SVM es encontrar el mejor hiper-plano que divida los datos de todas las muestras del entrenamiento en dos o más clases bien diferenciadas, es decir, determinar el hiper-plano con la máxima distancia con el punto de cada clase que está más cercano a éste. En la Figura 8 el hiper-plano H2 sería óptimo.

No obstante, cuando se utiliza SVM surgen dos problemáticas:

- Los universos que se estudian utilizan más de dos dimensiones y no tienen una representación lineal. Este problema se soluciona con la representación por funciones kernel, que proyectan la información a un espacio de características multidimensional mediante un mapeo no lineal.
- Seleccionar los parámetros apropiados del kernel. Hay dos parámetros en la función Función de Base Radial (RBF) del kernel (C y γ). Para encontrar los mejores parámetros de clasificación de prueba y entrenamiento se utiliza el método de optimización de parámetros.

IV-B. Algoritmo de Detección de Doble Compresión en Vídeos

Este algoritmo se utilizará con el fin forense de determinar si un vídeo ha sufrido más de una compresión, evidencia primera de que ese vídeo haya podido sufrir cualquier tipo de manipulación. La detección de recompresiones está basada

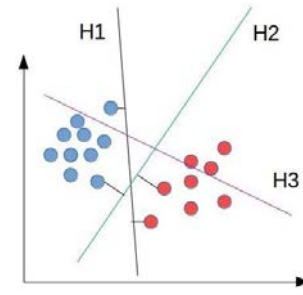


Figura 8: Muestras e hiperplanos.

en el estudio de las características estadísticas del Modo de Macrobloque (MBM) [27].

El MBM es una característica que consta del tipo de macrobloque y vector de movimiento. Para extraer esta característica, un vídeo es recomprimado repetidamente en la misma escala de calidad para luego calcular el número de MBM diferentes entre dos compresiones secuenciales. Finalmente, estas estadísticas extraídas son utilizadas por la SVM para determinar si el vídeo es original o si ha sido recomprimado.

Este método viene inspirado de la convergencia de los coeficientes Joint Photographic Experts Group (JPEG) cuando se recomprime una imagen. Ambos métodos para cada recompresión varían en la forma de la Figura 9.

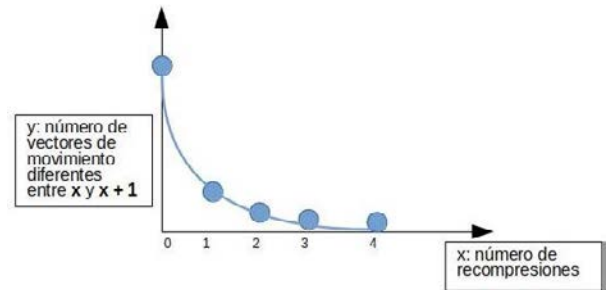


Figura 9: Número de MBM diferentes entre recompresiones.

A parte de los 3 tipos de fotogramas (I, P, B) que tiene el estándar MPEG4, también existen 3 tipos de macrobloques:

- I-MB: Macrobloques con intra-codificación.
- P-MB: Macrobloques con inter-codificación.
- S-MB: Macrobloques saltados.

Un MBM está compuesto de las dos propiedades de la siguiente manera:

$$[MBM(M) = M_{type}, M_{mv}]$$

$$M_{type} \in \{I - MB, P - MB, S - MB\}$$

$$M_{mv} = \{(u, v) | u, v \in Z\}$$

M es el macrobloque, M_{TYPE} el tipo del macrobloque M, y M-MV el vector de movimiento del macrobloque M.

Dos macrobloques se consideran que tienen el mismo MBM si y sólo si tienen el mismo M-TYPE y M-MV. Hay que tener en cuenta que cuando el M-TYPE es un I-MB su vector de movimiento es $\{0,0\}$, es decir, sólo es necesario evaluar los MBM diferentes de los P-fotogramas y por tanto sólo hay que comparar el Vector de Movimiento (VM).

Para una secuencia de compresiones sobre un vídeo, si el macrobloque del mismo fotograma y misma posición de

la compresión (n) y de la compresión (n + 1) tienen la característica MBM igual, se considera que ese macrobloque es estable. De otro modo se considera inestable. Ver Figura 10.

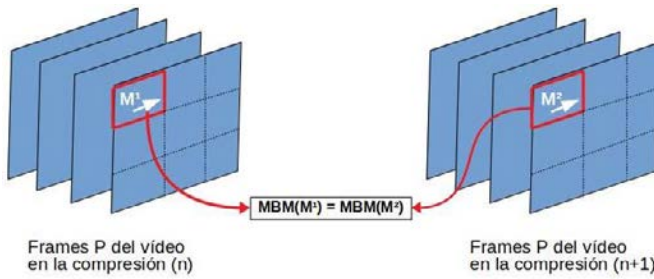


Figura 10: MBM estable.

A continuación se explica más detalladamente el algoritmo:

- Entrada: Vídeo o Vídeos en formato H.264-MP4
- Salida: Vector de características extraídas de los vídeos

Para un vídeo dado, este algoritmo devuelve un vector de características. En este caso el algoritmo se ha desarrollado para extraer tres características con el objetivo de que la SVM las pueda clasificar hasta la triple compresión. También es posible extraer sólo dos y así discernir sólo entre vídeos originales y doblemente comprimidos, o incluso, para detectar más allá de la triple compresión, aunque, a partir de ésta, la tasa de confianza de la máquina de soporte vectorial es demasiado bajo.

Vector de características:

- Número promedio de macrobloques inestables por P-fotograma encontrados entre el vídeo de entrada y su recompresión.
- Número promedio de macrobloques inestables por P-fotograma encontrados entre el vídeo recomprimido y su re-recompresión.
- Número promedio de macrobloques inestables por P-fotograma encontrados entre el vídeo de re-recompresión y su re-re-recompresión.

En primer lugar, se evalúa el MBM de los macrobloques. Para ello, con apoyo de la herramienta FFMPEG, se extraen los VM de los P-fotogramas. Es importante tener en cuenta que para poder extraer información el vídeo debe estar primero en un formato crudo (.YUV). Estos vectores de movimiento contienen la información del fotograma al que pertenecen, la posición del macrobloque y la posición del eje X, Y de origen

y destino del vector. Todos los vectores son almacenados en un archivo de texto para su posterior procesamiento.

Una vez realizado el proceso de extracción de VM, se debe recomprimir el vídeo de entrada con el formato H264-MP4. Esta recompresión tiene que tener la misma escala de calidad que el original, es decir la recompresión se realiza utilizando las mismas características del vídeo de entrada (qs, ancho, alto, rate, etc).

El vídeo recomprimido es almacenado para volver a ejecutar los pasos anteriores: Extracción de vectores de movimiento y, a partir de ahí, una nueva recompresión. Estas acciones pueden realizarse el número de veces que se quiera según el nivel de precisión de detección que se pretenda alcanzar. Para este caso se han realizado hasta tres recompresiones.

Cuando se alcance el número de recompresiones indicado, se va a disponer de tres archivos de texto que contienen los vectores de movimiento del vídeo de entrada y de sus recompresiones posteriores. Con toda esta información se puede proceder a calcular el número promedio de macrobloques inestables por P-fotograma (C).

La siguiente fórmula muestra cómo se realiza ese cálculo:

$$C_n = \frac{1}{N} \sum_{i,x,y} I(M_n(i, x, y), M_{n+1}(i, x, y))$$

N es el número total de P-fotogramas y M muestra el macrobloque de la recompresión enésima localizado en (x, y) del P-fotograma i-ésimo.

I se define como:

$$\left. \begin{array}{l} 1 \rightarrow MBM(M_1) \neq MBM(M_2) \\ 0 \rightarrow MBM(M_1) = MBM(M_2) \end{array} \right\} I(M_1, M_2)$$

Para realizar el cálculo de M(i,x,y), los vectores de movimiento contenidos en los ficheros de texto son tratados en forma de matrices N×M donde N es VM y M el P-fotograma al que pertenece con el fin de facilitar la tarea de la comparación.

Una vez encontrado I, es decir, el número de MBM diferentes de los ficheros de texto correspondientes a la compresión (n) y a la compresión (n + 1), se divide entre el número de P-fotogramas. El resultado se almacena en la posición (n) del vector de características del vídeo en cuestión.

Cuando el vector de características está completo se formatea para que la máquina de soporte vectorial lo pueda utilizar para realizar las tareas de clasificación. El diagrama de flujo de este algoritmo se muestra en la Figura 11.

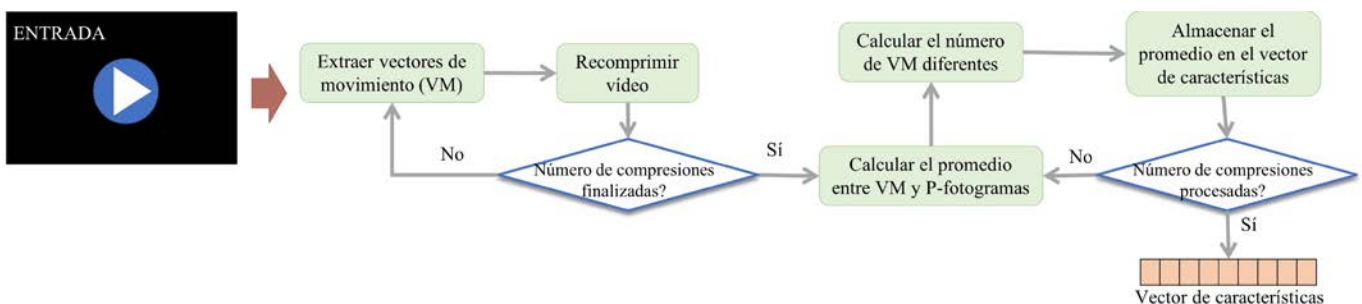


Figura 11: Diagrama del Algoritmo de Detección de Doble Compresión.

V. EVALUACIÓN DEL ALGORITMO PROPUESTO

Para evaluar el algoritmo propuesto se ha utilizado un dataset que contiene vídeos digitales procedentes de distintos modelos de dispositivos móviles con tamaños de resolución diferentes. Se han seleccionado vídeos en formato .MP4 con las resoluciones más comunes: 720x480, 720x1280, 1920x1080 y 3840x2160 (4K). La mayor parte de los vídeos seleccionados se han utilizado para el entrenamiento de la máquina de soporte vectorial con el fin de tener una base de conocimiento más amplia. El resto se han utilizado para la predicción. La Tabla I muestra un resumen de las características del dataset utilizado para el entrenamiento y predicción.

Tabla I: Dataset utilizado en la evaluación

Resolución	Formato	No. vídeos pruebas	No. vídeos analizados
720x480	MP4	20	20
720x1280	MP4	40	35
1920x1080	MP4	40	35
4K	MP4	20	17

Las características del equipo en el cual se han realizado los experimentos se presentan en la Tabla II. Es un factor importante a tener en cuenta ya que los tiempos de ejecución de las diferentes pruebas varían según los recursos computacionales disponibles.

Tabla II: Características del equipo de experimentación

Recursos	Características
Sistema operativo	Ubuntu 18.04
Memoria	4 GB
Procesador	Intel® Core™ 2 Quad CPU Q8200 @ 2.33GHz x 4
Gráficos	NV96
Tipo de SO	64 bits
Disco	100 GB

Para la realización de los experimentos se ha utilizado un componente llamado **MPEGflow** [33] bajo licencia Massachusetts Institute of Technology (MIT) que se apoya en FFMPEG para facilitar las tareas de extracción de los vectores de movimiento de los fotogramas de un vídeo y de almacenamiento en un fichero de texto.

También se ha utilizado el módulo LIBSVM, un software integrado que desempeña las funciones de una máquina de soporte vectorial, (C-SVC, nu-SVC), regresión (epsilon-SVR, nu-SVR) y estimación de distribución (SVM de una clase) compatible con la clasificación de clases múltiples.

A lo largo de esta sección se muestran todos los experimentos realizados para evaluar la efectividad del algoritmo de detección de manipulaciones basados en entrenamiento. Con ellos se pretende comprobar la variación de la precisión al aplicar el algoritmo sobre distintas resoluciones. Se estudia la capacidad de detectar si el vídeo es original, ha tenido doble compresión o triple compresión.

En primer lugar, se creó la base de conocimiento para el entrenamiento de la máquina de soporte vectorial. Los vídeos seleccionados como dataset de entrenamiento para cada resolución se utilizan como entrada del algoritmo de detección de recompresiones. Una vez entrenada la máquina con los vectores de características generados por el algoritmo, se puede comenzar la predicción. Esta prueba consiste en extraer las características de los vídeos a testear para que la máquina SVM, una vez entrenada, los clasifique en función de sus recompresiones.

En el primer grupo de experimentos tiene como objetivo detectar si un vídeo es original o ha tenido al menos una recompresión. En este caso, el algoritmo extrae sólo dos características de los vídeos del dataset. Por tanto, se entrena la máquina de soporte vectorial con vídeos originales y vídeos recomprimidos, el modelo resultante consta de dos clases para discernir si el vídeo es original o no. Se ejecutó un experimento para los vídeos con las diferentes resoluciones del dataset y adicionalmente se hizo un experimento mezclando todas las resoluciones para evaluar la tolerancia del algoritmo al tamaño del vídeo. Para cada uno de estos experimentos se han tomado los vectores de características escalados y sin escalar. Los resultados obtenidos para cada uno de los experimentos con diferentes resoluciones se presentan en la Tabla III.

Tabla III: Tasa de acierto para la detección de vídeos originales y manipulados con recompresión.

Tipo de datos	720x480	720x1280	Resolución 1920x1080	4K	MIX
Escalados	90 %	94,28 %	97,14 %	100 %	91,59 %
No escalados	90 %	100 %	100 %	100 %	92,52 %

Como se observa en la Tabla III, el algoritmo de detección propuesto presenta una tasa de acierto superior al 90 % incluso cuando la resolución es baja (720x480). En todas las resoluciones los resultados son superiores cuando los datos del vector de características no son escalados. Incluso en el caso de que el sistema sea entrenado con vídeos de diferentes resoluciones.

Las Tablas IV y V muestran las matrices de confusión resultantes para cada una de las resoluciones analizadas con los vectores de características escalados y sin escalar, respectivamente.

Tabla IV: Matriz de confusión por resolución con datos escalados.

Resolución	Matriz de Confusión				Acierto Prom.
	Clases	Original	Doble Comp.	Total Vídeos	
720x480	Original	10	0	10	90 %
	Doble Comp.	2	8	10	
720x1280	Original	20	0	20	94,28 %
	Doble Comp.	2	13	15	
1920x1080	Original	19	1	20	97,14 %
	Doble Comp.	0	15	15	
4K	Original	9	0	9	100 %
	Doble Comp.	0	8	8	
MIX	Original	44	9	53	91,59 %
	Doble Comp.	0	54	54	

Tabla V: Matriz de confusión por resolución con datos no escalados.

Resolución	Matriz de Confusión				Acierto Prom.
	Clases	Original	Doble Comp.	Total Vídeos	
720x480	Original	9	1	10	90 %
	Doble Comp.	1	9	10	
720x1280	Original	20	0	20	100 %
	Doble Comp.	2	0	15	
1920x1080	Original	20	1	20	100 %
	Doble Comp.	0	15	15	
4K	Original	9	0	9	100 %
	Doble Comp.	0	8	8	
MIX	Original	46	7	53	92,52 %
	Doble Comp.	1	53	54	

En el segundo grupo de experimentos se extrae una característica más que en los experimentos anteriores con el

objetivo de determinar si los vídeos analizados han sido recomprimidos más de una vez, y en tal caso, saber si ha sido recomprimido una o dos veces adicionales. Como en el caso anterior, para los experimentos se han tomado los vectores de características escalados y sin escalar. Los resultados obtenidos para cada uno de los experimentos con diferentes resoluciones se presentan en la Tabla VI.

Tabla VI: Tasa de acierto para la detección de vídeos con más de una recompresión.

Tipos de datos	Resolución				
	720x480	720x1280	1920x1080	4K	MIX
Escalados	83,33 %	70 %	68 %	80 %	49,67 %
No escalados	66,67 %	88 %	70 %	60 %	63,23 %

Como se observa en la Tabla VI, el algoritmo de detección propuesto presenta la mejor tasa de acierto (88 %) cuando el vídeo tiene una resolución de 720x1280 y los datos no escalados. Las Tablas VII y VIII muestran las matrices de confusión resultantes para cada una de las resoluciones analizadas con los vectores de características escalados y sin escalar, respectivamente.

Tabla VII: Matriz de confusión por resolución con tres clases con datos escalados.

Res.	Clases	Matriz de Confusión				Total Vídeos	Acierto Prom.
		Original	Doble	Triple			
720x480	Original	10	0	0		10	83,33 %
	Doble	0	5	5		10	
	Triple	0	0	10		10	
720x1280	Original	20	0	0		20	70 %
	Doble	6	9	0		15	
	Triple	1	8	6		15	
1920x1080	Original	20	0	0		20	68 %
	Doble	0	0	15		15	
	Triple	0	1	14		15	
4K	Original	9	0	0		9	80 %
	Doble	0	6	2		8	
	Triple	0	3	5		8	
MIX	Original	25	2	20		47	49,67 %
	Doble	0	0	54		54	
	Triple	1	1	52		54	

Tabla VIII: Matriz de confusión por resolución con tres clases con datos sin escalar.

Res.	Clases	Matriz de Confusión				Total Vídeos	Acierto Prom.
		Original	Doble	Triple			
720x480	Original	10	0	0		10	66,67 %
	Doble	1	5	4		10	
	Triple	0	5	5		10	
720x1280	Original	19	1	0		20	88 %
	Doble	0	12	3		15	
	Triple	0	2	13		15	
1920x1080	Original	20	0	0		20	70 %
	Doble	0	8	7		15	
	Triple	0	8	7		15	
4K	Original	7	2	0		9	60 %
	Doble	0	0	8		8	
	Triple	0	0	8		8	
MIX	Original	44	3	0		47	63,23 %
	Doble	6	29	19		54	
	Triple	7	22	25		54	

La Tabla IX muestra el tiempo de ejecución del algoritmo para cada una de las resoluciones.

La Tabla X muestra una comparativa entre el método propuesto y la investigación más relacionada en la literatura. En la tabla se observa que el resultado obtenido con el método propuesto para detectar doble compresión es superior al del trabajo comparado.

Tabla IX: Rendimiento del algoritmo propuesto.

Compresión detectada	720x480	720x1280	1920x1080	4K
Doble	00:00:07.03s	00:00:24.49s	00:01:16.32s	00:05:44.13s
Triple	00:00:16.23s	00:00:55.11s	00:02:37.32s	00:11:41.02s

Tabla X: Comparativa con la literatura.

Referencias	Características utilizadas	Dataset	Precisión	
			Doble Compresión	Triple Compresión
[27]	Macrobloques	YUV	94,10 %	–
Método propuesto	Vectores de Movimiento	Propio	95,27 %	75.33 %

VI. CONCLUSIONES

El contenido de imágenes y vídeos digitales posee información que va más allá de la visual. Esta información es de gran valor forense, pues su correcta explotación puede garantizar la autenticidad e integridad del contenido. Debido a esto, las imágenes y vídeos digitales son una excepcional fuente de evidencias a la hora de resolver procesos judiciales. El desarrollo y mejora continua de las nuevas tecnologías propicia que usuarios convencionales sean capaces de alterar el contenido de imágenes y vídeos con resultados profesionales, imperceptibles para el ojo humano. Ello se suma al hecho de que la detección de manipulaciones es una tarea compleja y también requiere de una mejora continua para adaptarse a tal escenario por lo que resulta imprescindible desarrollo de herramientas forenses capaces de detectar estas manipulaciones, cada vez más profesionales y habituales.

La línea de investigación que se ha seguido en este trabajo comienza realizando un estudio de las técnicas existentes de detección de manipulación sobre imágenes y vídeos digitales dedicando más esfuerzo a técnicas de detección de empalme en imágenes y en detección de doble compresión en vídeos.

Se ha diseñado e implementado una técnica de detección de manipulaciones basado en el estándar de vídeo H.264/MPEG4 para la detección de recompresiones en vídeos MP4 que compara los vectores de movimiento de los macrobloques de dos compresiones secuenciales del mismo vídeo para, a continuación, hacer uso de una SVM que clasifique el vídeo.

Se ha creado un dataset para evaluar la técnica de detección de recompresiones propuesta y se ha utilizado un dataset público para comparar los resultados con otras investigaciones relacionadas. La evaluación ha constado de dos experimentos divididos en grupos según la resolución de cada vídeo:

- Detección de vídeo original o doblemente comprimido, el algoritmo ha conseguido una precisión máxima con datos escalados del 100 % para vídeos de resolución 4K, para el resto de resoluciones no baja del 90 %.
- Detección de vídeo original, doble compresión, o triple compresión donde la precisión disminuye ligeramente respecto a la detección de original o doble compresión, tiene un promedio de precisión del orden del 80 %. No obstante, el mejor resultado lo sigue teniendo una alta resolución.

Los experimentos se han realizado tanto para datos sin escalar como para datos escalados, obteniendo unos resultados muy similares entre ellos. Por tanto no es relevante hacer un escalado de los mismos.

También se han realizado pruebas mezclando todas las resoluciones obteniendo unos resultados menos precisos que en aquellas pruebas donde sí se han separando las resoluciones. El rendimiento del algoritmo es directamente proporcional a la resolución del vídeo que se quiera procesar y a la cantidad de recompresiones que se quieran detectar.

En base a los resultados obtenidos en esta investigación, las líneas futuras de investigación que se proponen en el presente trabajo son las siguientes:

- Extender el algoritmo de detección de recompresiones para utilizarlo con otros códecs de vídeo a parte del H264.
- Utilizar técnicas de aprendizaje profundo y aumentar el número de características extraídas para mejorar la precisión de detección de recompresiones.
- Optimizar el algoritmo de detección de recompresiones para reducir el tiempo de procesado en vídeos de alta resolución.

AKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700326. Website: <http://ramses2020.eu>. This paper has also received funding from THEIA (Techniques for Integrity and authentication of multimedia files of mobile devices) UCM project (FEI-EU-19-04).



REFERENCIAS

- [1] E. M. Nieto, "The value of photography: Anthropology and image."
- [2] S. J. Nightingale, K. A. Wade, and D. G. Watson, "Can people identify original and manipulated photos of real-world scenes?" *Cognitive research: principles and implications*, vol. 2, no. 1, p. 30, 2017.
- [3] A. De, H. Chadha, and S. Gupta, "Detection of forgery in digital video," in *The 10th World Multi Conference on Systemics Cybernetics and Informatics*, vol. 5, 2006, pp. 229–233.
- [4] W. Wang and H. Farid, "Exposing digital forgeries in interlaced and deinterlaced video," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 438–449, 2007.
- [5] N. Mondaini, R. Caldelli, A. Piva, M. Barni, and V. Cappellini, "Detection of malevolent changes in digital video for forensic applications," in *Security, steganography, and watermarking of multimedia contents IX*, vol. 6505. International Society for Optics and Photonics, 2007, p. 65050T.
- [6] M. Kobayashi, T. Okabe, and Y. Sato, "Detecting forgery from static-scene video based on inconsistency in noise level functions," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 883–892, 2010.
- [7] P. Bestagini, S. Battaglia, S. Milani, M. Tagliasacchi, and S. Tubaro, "Detection of temporal interpolation in video sequences," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3033–3037.
- [8] Y. Yao, G. Yang, X. Sun, and L. Li, "Detecting video frame-rate up-conversion based on periodic properties of edge-intensity," *Journal of Information Security and Applications*, vol. 26, pp. 39–50, 2016.
- [9] M. Xia, G. Yang, L. Li, R. Li, and X. Sun, "Detecting video frame rate up-conversion based on frame-level analysis of average texture variation," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 8399–8421, 2017.
- [10] G. Chetty, "Blind and passive digital video tamper detection based on multimodal fusion," in *Proc. of the 14th WSEAS International Conference on Communications*, 2010, pp. 109–117.
- [11] J. Goodwin and G. Chetty, "Blind video tamper detection based on fusion of source features," in *Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on*. IEEE, 2011, pp. 608–613.
- [12] S. Das, G. Darsan, L. Shreyas, and D. Devan, "Blind detection method for video inpainting forgery," *International Journal of Computer Applications*, vol. 60, no. 11, 2012.
- [13] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, "Local tampering detection in video sequences," in *Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on*. IEEE, 2013, pp. 488–493.
- [14] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting duplication," in *Proceedings of the 9th workshop on Multimedia & security*. ACM, 2007, pp. 35–42.
- [15] C.-S. Lin and J.-J. Tsay, "A passive approach for effective detection and localization of region-level video forgery with spatio-temporal coherence analysis," *Digital Investigation*, vol. 11, no. 2, pp. 120–140, 2014.
- [16] R. C. Pandey, S. K. Singh, K. Shukla, and R. Agrawal, "Fast and robust passive copy-move forgery detection using surf and sift image features," in *Industrial and Information Systems (ICIIS), 2014 9th International Conference on*. IEEE, 2014, pp. 1–6.
- [17] D.-K. Hyun, S.-J. Ryu, H.-Y. Lee, and H.-K. Lee, "Detection of upscale-crop and partial manipulation in surveillance video based on sensor pattern noise," *Sensors*, vol. 13, no. 9, pp. 12 605–12 631, 2013.
- [18] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting double mpeg compression," in *Proceedings of the 8th workshop on Multimedia and security*. ACM, 2006, pp. 37–47.
- [19] —, "Exposing digital forgeries in video by detecting double quantization," in *Proceedings of the 11th ACM workshop on Multimedia and security*. ACM, 2009, pp. 39–48.
- [20] Y. Su and J. Xu, "Detection of double-compression in mpeg-2 videos," in *Intelligent Systems and Applications (ISA), 2010 2nd International Workshop on*. IEEE, 2010, pp. 1–4.
- [21] Y. Su, W. Nie, and C. Zhang, "A frame tampering detection algorithm for mpeg videos," in *Information Technology and Artificial Intelligence Conference (ITAIC), 2011 6th IEEE Joint International*, vol. 2. IEEE, 2011, pp. 461–464.
- [22] T. Sun, W. Wang, and X. Jiang, "Exposing video forgeries by detecting mpeg double compression," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1389–1392.
- [23] J. Xu, Y. Su, and X. You, "Detection of video transcoding for digital forensics," in *Audio, Language and Image Processing (ICALIP), 2012 International Conference on*. IEEE, 2012, pp. 160–164.
- [24] A. Gironi, M. Fontani, T. Bianchi, A. Piva, and M. Barni, "A video forensic technique for detecting frame deletion and insertion," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 6226–6230.
- [25] A. W. A. Wahab, M. A. Bagiwa, M. Y. I. Idris, S. Khan, Z. Razak, and M. R. K. Ariffin, "Passive video forgery detection techniques: a survey," in *Information assurance and security (IAS), 2014 10th International Conference on*. IEEE, 2014, pp. 29–34.
- [26] X. Jiang, W. Wang, T. Sun, Y. Shi, and S. Wang, "Detection of double compression in mpeg-4 videos based on markov statistics," vol. 20, pp. 447–450, 05 2013.
- [27] J. Chen, X. Jiang, T. Sun, P. He, and S. Wang, "Detecting double mpeg compression with the same quantiser scale based on mbm feature," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 2064–2068.
- [28] X. Jiang, P. He, T. Sun, F. Xie, and S. Wang, "Detection of double compression with the same coding parameters based on quality degradation mechanism analysis," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 1, pp. 170–185, 2018.
- [29] J. A. Aghamaleki and A. Behrad, "Detecting double compressed mpeg videos with the same quantization matrix and synchronized group of pictures structure," *Journal of Electronic Imaging*, vol. 27, no. 1, p. 013031, 2018.
- [30] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h. 264/avc video coding standard," *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 560–576, 2003.
- [31] "Ffmpeg," <https://www.ffmpeg.org>.
- [32] "Debug/macroblocksandmotionvectors," https://trac.ffmpeg.org/attachment/wiki/Debug/MacroblocksAndMotionVectors/vismv_pf.png.
- [33] "github: vadimkantorov/mpegflow," <https://github.com/vadimkantorov/mpegflow>.

Edgar González Fernández was born in Mexico City. He received a Degree in Applied Mathematics from the Universidad Autónoma del Estado de Hidalgo in 2010, and a Master in Science with Specialization in Mathematics from the Center for Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV-IPN). He is currently a Ph.D. student in the Computer Science Department at CINVESTAV-IPN. Currently he is Member in the Research Group GASS (Group of Analysis, Security and Systems, <http://gass.ucm.es>) at the Universidad Complutense de Madrid (UCM). His research interests are Cryptography, Information Security and Data Science.

Ana Lucila Sandoval Orozco was born in Chivolo, Magdalena, Colombia in 1976. She received a Computer Science Engineering degree from the Universidad Autónoma del Caribe (Colombia) in 2001. She holds a Specialization Course in Computer Networks (2006) from the Universidad del Norte (Colombia), and holds a M.Sc. in Research in Computer Science (2009) and a Ph.D. in Computer Science (2014), both from the Universidad Complutense de Madrid (Spain). She is currently a postdoctoral researcher

at Universidad Complutense de Madrid (Spain). Her main research interests are coding theory, information security and its applications.

Luis Javier García Villalba received a Telecommunication Engineering degree from the Universidad de Málaga (Spain) in 1993 and holds a M.Sc. in Computer Networks (1996) and a Ph.D. in Computer Science (1999), both from the Universidad Politécnica de Madrid (Spain). Visiting Scholar at COSIC (Computer Security and Industrial Cryptography, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, Belgium) in 2000 and Visiting Scientist at IBM Research Division (IBM Almaden Research Center, San Jose, CA, USA) in 2001 and 2002, he is currently Associate Professor of the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM) and Head of Complutense Research Group GASS (Group of Analysis, Security and Systems) which is located in the School of Computer Science at the UCM Campus. His professional experience includes research projects with Hitachi, IBM, Nokia and Safelayer Secure Communications.

TENDENCIAS EN SEGURIDAD

Methodology to Securitize Smart Toys in Household Environments

P. González, F. Paniagua, D. Suárez and J. J. Nombela

Resumen — This paper examines the present situation regarding Smart Toys, the privacy and security concerns that they generate for children and their relatives, as well as the existing solutions that would allow them to improve the security of these devices. On this purpose, it is examined the state of art in this field, including some recent data breaches concerning Smart Toys, and the present available solutions to identify and mitigate their security risks. As a result of not identifying a suitable methodology that provides a unified approach to analyze and securitize household environments composed by different Smart Toys that can easily be implemented by non-technical users, in this paper is proposed a new and tailored methodology. The presented proposal aims to tackle and solve the identified security concerns and present the information in a user-friendly manner, helping final users to understand and address the security issues of their Smart Toys, even without having a deep technical knowledge in the field.

Palabras clave — Methodology, Security and Privacy Protection, Risk Management, Internet of Things, Smart Toys.

I. INTRODUCTION

THE NUMBER of IoT devices and services which are being used nowadays has been experimenting an exponential growth for the past ten years. According to Gartner's predictions, it is expected that by 2020 the number of connected devices will reach the 20 billion [1], and other sources, such as Mozilla, share even more ambitious figures [2].

Along with the increase of the number of devices in use there is also an important expansion of the vulnerability surface for these devices and the number of threats they are exposed to. This situation is also concerning to Smart Toys users, which showed on an ESET survey that more than one third of them are very worried about the privacy and security of children using these devices [3].

Bearing this in mind, it was conducted a deep research on the existing solutions that could allow users to identify the threats and risks surrounding the Smart Toys they possess, as well as the security measures that could be implemented to mitigate the identified risks.

As a result, several security risk methodologies, frameworks and guidelines were analyzed, but most of them were hardly adapted to such a specific IoT environment. On top, most of the identified resources were targeted to an

audience with a reasonable level of technological understanding, which is why they could not be implemented by the average users of Smart Toys (children and/or parents).

These circumstances lead to a complex situation: the users of Smart Toys not only ignore the risks they are exposing their families to, but in case of having interest in securitizing their IoT environment, they cannot rely on a user-friendly procedure or reference that helps them to accomplish this mission [4].

Thus, this paper analyzes the present situation of Smart Toys, the threats that can affect them, and the methodologies that could be used as a reference to securitize IoT environments. Also, and due to the deficiencies of the existing methodologies to solve the aforementioned problems, it is proposed a methodology that allows users to identify, understand and take decisions about the security settings of their Smart Toys.

The methodology proposed is also validated, using a real case of a Smart Toy data breach, to verify that it can actually contribute in identifying the risks these toys can expose users to, improving their knowledge and understanding of them. Moreover, it can also be used as a reference for developing a security framework that can be used to identify risks and mitigations of household IoT devices, or even other kind of environments with similar features.

The rest of the paper is organized as follows. Section 2 presents briefly an overview of the IoT architectures and security. Section 3 analyzes the security gaps for Smart Toys. Section 4 describes the proposed methodology. Section 5 tests and validates the proposed methodology. Section 6 presents some conclusions extracted from the work carried out in this paper.

II. CONTEXT AND STATE OF THE ART

Definition of Smart Toys

To understand what a Smart Toy makes reference to, it is important to first understand what the concept IoT (Internet of Things) stands for.

Even though there is not an official definition of the term yet, if we consider the definitions provided by reference institutions such as the IEEE [5], ENISA [6] or Gartner [7], there is certain consensus agreeing that the term IoT comprises a wide ecosystem of interconnected services and devices.

The Smart Toys are just a specific type of IoT that are intended to interact with children and their environment as part of a leisure activity.

P. González, Universidad Internacional de La Rioja, Madrid, Spain, paulagonzalezdom@gmail.com

F. Paniagua, Universidad Carlos III, Madrid, Spain, fidel.paniagua@uc3m.es

D. Suárez, Universidad Internacional de La Rioja, Madrid, Spain, diego.suarez@unir.net

J. J. Nombela, Universidad Internacional de La Rioja, Madrid, Spain, juanjose.nombela@unir.net

Corresponding author: Fidel Paniagua Díez

IoT Architectures

Likewise it happens with the definition of the term, there is not a standard model or architecture for IoT environments, and the proposals varie depending on the source consulted.

In broad strokes the architecture of an IoT environment can be represented by three, four or five layers [8], to understand the main interactions among all the involved devices:

- (i) *Three Layer Architecture*: Differentiates between the Perception Layer (to identify and collect information), the Network Layer (to transmit data between layers) and the Application Layer (defines the applications related to the IoT).
- (ii) *Four Layers Architecture*: Apart from the previous layers, it also considers a Support Layer, which contains the security implementations.
- (iii) *Five Layers Architecture*: Considers the three first layers, named the Perception, Transport and Application layers, and two more layers, which are the Processing Layer (middleware layer that processes the collected information) and the Business Layer (allows the management of the whole IoT system).

Apart from the generic architectures mentioned above, there are further architecture models that renowned institutions, such as Gartner, propose, such as their Reference Model composed by five layers (Process, Function, Information, Communication and Device) and three tiers (Edge, Platform and Enterprise) [7].

Available Security Solutions

As it has been shown, understanding and analyzing IoT environments can be a challenging task, due to the general lack of consensous. This is an important handicap when the Smart Toys, or IoT in general, are studied from a security point of view.

Nowadays, there are several institutions working towards improving the security of IoT environments. Concerning children data privacy there are two main regulations around the world that are having a great impact in this context:

- (i) *COPPA (Children's Online Privacy Protection Act)*: In the United States of America, this regulation defines certain requirements that operators of websites and online services must consider when their services are directed to children under 13 or when they are collecting personal information of them [9].
- (ii) *GDPR (General Data Protection Regulation)*: This regulation defines a common data protection frame in Europe, imposing great sanctions to data processors and controllers that do not use or are not implementing suitable security measures for the data they are processing [10].

There are also institutions that, even if they do not have a legal influence over the design or production Smart Toys, are developing resources such as guides, methodologies or frameworks to improve the overall security of IoT environments; some of them have even make some contributions to the specific field of Smart Toys, providing generic guidelines showing the basic steps to improve the security of a Smart Toy [11]. However, the overall maturity level of the field, in terms of standardization, is still low.

Some of the most representative contributions in this field are the following ones:

- (i) *OWASP IoT Project*: In 2015 the Open Web Application Security Project presented a draft about the attack surface areas and security considerations for IoTs [12].
- (ii) *European Union Agency for Network and Information Security (ENISA)*: In 2017 ENISA published a study titled '*Baseline Security Recommendations for Internet of Things in the context of critical information infrastructures*', which aims to set the scene for IoT security in Europe, providing insight into the security requirements of IoT, mapping critical assets and relevant threats, assessing possible attacks and identifying potential good practices and security measures to apply in order to protect IoT systems. It serves as a reference point in this field and as a foundation for relevant forthcoming initiatives and developments [6].
- (iii) *National Institute of Standards and Technology (NIST)*: In February of 2018 the '*NIST Internal Report*' (NISTIR) published a draft for IoT security, that aims to inform and enable policymakers, managers, and standards participants as they seek timely development of and use of cybersecurity standards in IoT components, systems, and services. It brings a good frame to understand the cybersecurity landscape for IoT and identify the areas where security standards are missing, but it does not provide yet a security framework itself [13].
- (iv) *Gartner*: In 2018, the Gartner Advisory Company released the already mentioned paper called '*Architect IoT Using the Gartner Reference Model*' which provides an architecture blueprint that defines what functionality is required, where that functionality will operate, and how data and control will flow in an IoT project [7].
- (v) *Microsoft*: In 2018 Microsoft Azure published a Security Architecture for IoTs that through a Threat Model defines a path to identify security threats in very diverse contexts [14].

Even though all the abovementioned efforts contribute with very interesting insights into the field of study, there are still important security issues that are not addressed or solved by these proposals, leaving children and their relatives totally exposed to data breaches and other security and safety risks of Smart Toys.

III. SECURITY ANALYSIS

Security Breaches for Smart Toys

Within the past years, there have been several security breaches concerning Smart Toys which have compromised the Privacy and Security of the children using it, as well as their own Safety and of the ones surrounding them. The much talked-about are the following:

- (i) *VTech*: The company VTech sells toys and gadgets for children, including tablets, phones and baby monitors. In November 2015, it was announced that a security breach had occurred and that it exposed personal information and photos of almost 5 million parents and more than 6 million kids due to bad security practices (weak encryption algorithms or lack of them, passwords stored in plain text, vulnerabilities to SQL injections, etc) [15].

(ii) *CloudPets*: The Smart Toy allowed children to send and receive audio recordings between them and an external app that can be installed in most of the smartphones and tablets of the children's relatives. At the beginning of 2017 an important leakage of 820,000 user accounts was reported, which included the personal information, photos and recordings mainly of children, but also of their families, mainly on account of the data storing and the production website were publicly facing a network segment without any authentication or password required, apart from other security misconfigurations [16] [17].

Apart from the above mentioned ones, there are some other examples that also show not only the consequences to Smart Toy security breaches, but also the risk potential that these devices have, as well as the general perception of customers about them. Some of the following are very representative:

(i) *Hello Barbie*: By the end of 2015, the toy company Mattel and the technology company ToyTalk announced the launch of a Smart Barbie called '*Hello Barbie*'. Apart from collecting data, the Hello Barbie owed an Artificial Intelligence (AI) and a voice recognition software that allowed the toy to have very realistic conversations with the children and adapt her answers. Many concerns were raised when the company stated that the voice-recordings would be shared with third parties to improve the experience of the users, which lead to a public petition with more than 37,000 signatures to drop the toy from the market [18].

(ii) *My Friend Cayla and i-Que Robot*: Another two Smart Toys based on AI called '*My Friend Cayla*' and '*i-Que Robot*', launched by the company Genesis Toys in collaboration with ToyQuest and Nuance Communications, had a very bad reception from the parents and even federal institutions on their latest release by the end of 2016. The allegations were based on the poor security features of the doll and the position of the company about sharing data with third parties [19].

Apart from the present field of study (Smart Toys), there are other IoT devices that can interact with these devices or their targeted users, such as Echo devices, which have also brought to attention security concerns that should be addressed. A well known example is the case of a six years old kid who bought a dollhouse just by ordering it to the Echo device [20].

Gaps in the Existing Security Solutions

The just-mentioned security breaches have occurred and impacted data subjects despite the previously mentioned existing security solutions. This happened because of the also mentioned security gaps surrounding these solutions. As a summary, can be pointed out the following three main problems:

(i) *Lack of agreement on the considerations*: The lack of consensus concerning all the IoT world generates a lot of doubts about which procedure to follow to identify security risks and mitigations for Smart Toys. Moreover, some of the mentioned methodologies are too specific for the environments they have been designed for, not allowing the flexibility of implementing them for this specific use case.

(ii) *The implementation procedure is not clear*: Aligned with the previous point, even though the regulations and frameworks help to promote a security culture among Smart Toy manufacturers, they refer to what should be considered but not how to implement their requirements on products and services.

(iii) *Language and approach*: Most of the available solutions are oriented to a target audience that can interfere at the first stages of the commercialization of a Smart Toy (suchs as manufacturers or intermediates), which most likely will have certain technical knowledge. However, when these agents fail to provide the pertinent security to their products and services, final users are very vulnerable to security breaches. Most of them are not aware of the security risks their use can imply, or, even in the case of being interested, the available sources or information are not addressed in a user-friendly language, hindering their ability to interfere in the management of the IoT security.

Apart from these main problems, there are also other additional considerations that should not be forgotten, such as the diversity of features and elements that can comprise each Smart Toy or IoT system (different communication protocols, software features, etc.), or the wide assortment of vendors, functions and outcomes for every new toy that is introduced in the market.

IV. METHODOLOGY PROPOSAL

This paper proposes a modular security methodology that can be used in different household environments with different Smart Toys. The principal focus is to cover the security gaps previously identified and to develop a security methodology that allows the identification of Smart Toys security risks and mitigations, in a flexible, but still complete way.

It consists on a set of six defined steps which are the following:

- (i) *First Step*: Delimitation of the Environment
- (ii) *Second Step*: Identification of the Roles Involved
- (iii) *Third Step*: Identification of the Technologies Involved
- (iv) *Fourth Step*: Identification of the Functions for each Technology
- (v) *Fifth Step*: Identification of the Security Risks for each Function
- (vi) *Sixth Step*: Identification of Mitigations for each Risk

Each one of these steps will be explained in detail in the following sections.

First Step: Delimitation of the Environment

The first step consists in the definition and delimitation of the physical and virtual space that composes the Smart Toy environment.

Using as a reference the already mentioned architectures and methodology sources, in this paper is proposed a simplified architecture for a household setting, to contextualize the environment where the Smart Toys are located. The main elements contained would be the following:

(i) *Edge Devices (Smart Toys)*: It refers to the virtual and physical elements that integrate the IoT ecosystem.

(ii) *Edge Computing*: It refers to a distributed IT architecture, in which user data are collected, stored, exchanged and processed at the periphery of the network, but still close to the original source of the data. This allows the processing of time-sensitive data in almost real time, avoiding also the time lapse and costs derived from Cloud Computing.

(iii) *Edge Gateway*: It refers to the physical or virtual node that serves as the connection point between different Edge Devices, as well as between the IoT ecosystem and outsider networks. It provides system interoperability, communication and data-processing capabilities, among other features.

(iv) *Cloud Computing*: It refers to the use of remote services such as software, platforms or infrastructure, to store, process and retrieve data from an off-site location. In the IoT context it is generally used for historical analysis, big data analytics and long-term storage.

(v) *Cloud Backend*: It refers to the server side on a Cloud Computing service where all the processes actually take place.

As it is shown, the environment comprises not only the physical space where the devices are allocated, but also the area of influence of the communications among devices and their connections with external parties. On this purpose, it is important to list the elements present in the space but also to represent them graphically to better understand their connections and interactions.

Second Step: Identification of the Roles Involved

The second step consists in the identification of all the roles that can interact or influence the IoT ecosystem. The role is a representation of one or several natural or legal people, that share some common particularities.

Considering the kind of actors that can interact with the Smart Toy environment, in this paper are proposed three main categories of roles with similar features:

(i) *Users*: Commonly with low cybersecurity knowledge, access to security solutions and almost no influence on the technical design and features of the devices

(ii) *Authorized third parties*: such as the producers or service providers, which commonly prioritize commercial features instead of security ones.

(iii) *Unauthorized third parties*: such as intruders or unintended natural or legal persons which can interact with the Smart Toys.

This classification is important to understand who can influence the security of the IoT ecosystem, either through the improvement of it or through the generation of risky situations.

Therefore, depending on the focus of interest while implementing this methodology, it will be more or less interesting to also include certain level of granularity on the classification, instead of using the three previously-defined categories of roles for a household environment. The level of granularity can especially be affected depending on who is implementing the methodology, and which are the specific purposed for its implementation.

Third Step: Identification of the Technologies Involved

The third step consists in the identification of all the technologies involved in the IoT ecosystem.

Considering all the sources mentioned in the research analysis, but in particular ENISA [6] and Gartner [7], in this paper is proposed the following classification:

(i) The main devices should have the the following IoT features:

- *Tags* (to identify the device)
- *Sensors* (capacity that allows to collect data)
- *Communicators* (to transfer the data)
- *Actuators* (to take actions depending on the processing of the data)
- *Software/Hardware* (to be able to perform and support the processing operations)

(ii) Peripheral elements that can interact with the main device:

- *Physical devices* (e.g. remote controls)
- *Software elements* (e.g. applications)
- *Elements of the IoT network* (e.g. routers or gateways)

This identification and classification should be assigned to each of the IoT items composing the IoT environment.

Fourth Step: Identification of the Functions for each Technology

The fourth step consists in the identification of all the functions that each of the previously identified technologies possess.

Based on the existing researches in the field, in this paper are proposed a group of six main functions that a Smart Toy could perform, regarding to data processing:

(i) *Data Collection*: This function is carried out mainly by the sensors. They can detect changes in a physical or virtual level.

(ii) *Data Storing*: The data gathered by the sensors or provided by other IoT elements can be stored temporarily or permanently at the edge and/or at the cloud.

(iii) *Data Analysis*: The data gathered can be manipulated in order to obtain information that can be used by the IoT and provide a specific output in a timely manner. Common manipulations are the aggregation, organization, transformation or even deletion of data, to allow a faster and more efficient processing of the data gathered.

(iv) *Data Transmission*: The data can be transmitted to peripheral elements using wired or wireless communication technologies.

(v) *Data Display*: The data gathered and processed allows the IoT actuators and interfaces to show or display specific behaviors, that can cause a physical and/or a digital impact.

(vi) *Data Management*: This concept comprises the ability of influencing data to take decisions over all the previous functions, as well as about aspects such as the purpose of the data processing, the access and security of this data, among other functional decisions.

To be able to trace the steps, it is recommended to design a table, a tree diagram, or a similar structure, that allows to link each function to each of the identified technology, while it also provides a visual representation of where the most sensitive areas of the IoT can be located.

Fifth Step: Identification of the Security Risks for each Function

The fifth step consists in the identification of security risks associated to each of the functions previously defined, as well as which of the studied roles can cause each risk.

Consequently, at this point is necessary not to only identify the set of risks affecting the Smart Toy environment, but also to correlate each of them with the roles identified in previous steps. It is suggested to continue with the representation previously chosen (table, chart, etc.), allowing an easy correlation between the function of each technology and its correspondent risk.

Regarding to the correlation of each risk with the role generating it, would be as simple as to include another section (column, brunch, etc.) in the chosen representation, where this role is indicated following the previous classification. In Table 1 are described the most common correlations:

TABLE 1
RISK-ROLES CORRELATION

ROLE	R/R ID	ORIGIN
Provider	PR1	Poor security design of the IoT architecture
	PR2	Misuse of data from this party
User	UR1	Poor choice of the security configurations or actions
	UR2	Misuse of the device or system

As a consequence, unauthorized third parties could intrude or affect the IoT systems endangering their confidentiality, integrity or availability and its related roles.

In the Table 2 has been developed a high level classification of security risks that can affect IoT environments.

TABLE 2
RISK-FUNCTIONS CORRELATION

FUNCTIONS	RISK ID	DEFINITION
Data Collection	CR1	The information is not collected or is not accurate
	CR2	Unauthorized information is collected
Data Storing	SR1	If the databases are not secured, the collected data can be accessed or tampered
	SR2	If there is no backup of the information, it can be lost
Data Analysis	AR1	If the software is tampered or misconfigured, the results of the data analysis could not be the intended ones
	AR2	If the data provided is not essential to carry out the main activity of the device, could be used for illegitimate purposes

	AR3	Depending on the data provided, could be used for profiling
Data Transmission	TR1	If someone is located in the transmission area, can eavesdrop or tamper the data collected or displayed
	TR2	If the transmission is not encrypted, it can be eavesdropped or tampered
	TR3	If someone interferes the connection, can eavesdrop or tamper the data collected or displayed Data Display
Data Display	DR1	Legitimate information is not displayed
	DR2	Tampered information is displayed
Data Management	MR1	If there is no access control or it is inadequate, unauthorized users can access the device and its features
	MR2	If there is no authentication control or it is inadequate, unauthorized users can impersonate legitimate ones
	MR3	If there are no lost access procedures, or they are inadequate, legitimate users could lose access to the services provided
	MR4	If there are no update measures or they are inadequate, security vulnerabilities could not be patched
	MR5	If the password policies are not strong enough, unauthorized users can impersonate legitimate ones
	MR6	If the application has not been securely developed, data can be accessed or tampered

Even though the list provided is not exhaustive, in conjunction with the rest of identified elements, can provide a clear picture about the origin and impact of the security risks. Moreover, this classification can also be complemented with Information Security generic risks, which can be obtained from risks or threats catalogues such as the ENISA Threat Taxonomy [6].

Sixth Step: Identification of Mitigations for each Risk

The sixth step consists in the identification of mitigations that can be implemented to reduce or even eliminate the risks associated to the functions of each technology that composes the Smart Toys ecosystem, as well as which of the studied roles can mitigate that risk.

Consequently, as in the previous step, it is necessary to identify the set of mitigations for each identified risk, and to correlate it with the role that can implement the suggested mitigation. Likewise, it is suggested to continue with the previously chosen representation (table, chart, etc.) and proceed like it is described on the step 5.

Regarding to the correlation of each mitigation with the role generating it, the most common correlations would be the ones described in Table 3:

TABLE 3
MITIGATION-ROLES CORRELATION

ROLE	R/M ID	MITIGATION
Provider	PM1	Can only be implemented by design and before the IoT is distributed
	PM2	Can be implemented remotely and once the IoT is already operating
User	UM1	Can alter the security configurations of the IoT to mitigate the risk
	UM2	Cannot mitigate the risk, meaning that it is necessary to decide whether accepting the risk or rejecting the use of the IoT

In the same way that it happened in the previous step, the identification of security mitigations is also too broad to be tackled in the present project. For this reason, in Table 4 are only represented the main identified mitigations for the previously described risks. Moreover, the technical implementation of each of these mitigations will also depend on the choice of the implementor and the available resources for this purpose.

TABLE 4
RISK-MITIGATIONS CORRELATION

FUNCTION	RISK ID	MIT. ID	MITIGATION DESCRIPTION
Data Collection	CR1	CM1.1	To include an informing feature when the information is properly collected
		CM1.2	To include integrity solutions, such as hashing or certificates
	CR2	CM2	To include an information-restricted configuration, attending to the principle of Need-to-know.
Data Storing	SR1	SM1	If the storing service is outsourced, only use trusted vendors and solutions
	SR2	SM2	To define backup solutions or infrastructures
Data Analysis	AR1	AM1	To require and provide only the essential information to carry out the expected service.
	AR2	AM2	To require and provide only the essential information to carry out the expected service.
	AR3	AM3	To require and provide

			only the essential information to carry out the expected service.
Data Transmission	TR1	TM1.1	To monitor all the connections established
		TM1.2	To activate diode functionalities (only entrance/only release)
	TR2	TM2.1	To enable and use secure transmission protocols
		TM2.2	To include integrity solutions, such as hashing or certificates
	TR3	TM3.1	To monitor all the connections established
		TM3.2	To enable and use secure transmission protocols
		TM3.3	To close unused ports
		TM3.4	To activate diode functionalities (only entrance/only release)
Data Display	DR1	DM1	To include an informing feature when the information is displayed, or it fails to be displayed
	DR2	DM2	To include integrity solutions, such as hashing or certificates
Data Management	MR1	MM1.1	To include access control features
		MM1.2	To use the principle of Need-to-know
	MR2	MM2.1	To provide solid authentication control features
		MM2.2	To use hardened passwords, codes or methods of authentication to avoid impersonation based on easy-to-access private data of the user
	MR3	MM3.1	To use hardened lost access procedures control
		MM3.2	To avoid the use of unreasonable requirements for access control in non-critical cases, that would lead the user to easily lose or forget the access credentials
	MR4	MM4.1	To provide means of update for devices and services
		MM4.2	To verify the software keeps updated with the last released versions
	MR5	MM5.1	To guarantee hardened password policies
		MM5.2	To avoid unsecure

			practices such as the reuse of passwords or the use of data easy to guess or to obtain
	MR6	MM6.1	To guarantee S-SDLC practices
		MM6.2	To only use applications and devices provided by trusted vendors

The provided list can also be used as a base for future studies, where it can also be completed and updated accordingly to the evolution of the currently studied IoT field.

V. VALIDATION OF THE METHODOLOGY

To test if the proposed methodology can successfully achieve the goal it has been designed for, it is layed out a real scenario with the CloudPet Smart Toy. This has been the toy of choice due to the great number of vulnerabilities identified that lead to its previously mentioned data breach [16 [17].

The proposed methodology defines a process step by step which allows to identify and map the features of the toy with all the described elements, resu lting in the identification of risks for the toy, as well as the roles generating them. The output of it is shown in Table 5.

TABLE 5
CLOUDPETS RISKS IDENTIFICATIONS

ASSETS	TECHN.	FUNCTIONS	RISKS	ROLES
Main Device	Sound Processor	D. Display	DR2	PR1
	Bluetooth LE	D. Transmission	TR1, TR2, TR3	PR1, UR2
	Wi-Fi	D. Transmission	TR2, TR3	PR1, UR1, UR2
Periph. Elements	Web App	D. Analysis	AR2	PR1, PR2
		D. Storing	SR1	PR1
		D. Management	MR, MR4, MR6	PR1
			MR2	UR2
			MR5	PR1, UR1
	Mobile App	D. Management	MR1, MR4, MR6	PR1
			MR2	UR1
			MR5	PR1, UR1
		D. Collection	CR2	PR1, UR1, UR2
		D. Display	DR2	PR1

Considering the risks identified by the methodology and the ones pointed out by prestigious security professionals such as Troy Hunt [16]or Paul Stone [17], it can be seen that they are actually aligned.

Once the risks have been identified, it is also possible to identify the security mitigations that would be related to each of the identified risks, as it is shown in Table 6.

TABLE 6
CLOUDPETS RISK MITIGATIONS

RISKS ID	MIT. ID	IDENTIFIED MITIGATION	ROLES
SR1	SM1	If the storing service is outsourced, only use trusted vendors and solutions	PM1, PM2, UM1
AR2	AM2	To require and provide only the essential information to carry out the expected service.	PM1, PM2, UM1
TR1	TM1.1	To monitor all the connections established	UM1
	TM1.2	To activate diode functionalities (only entrance/only release)	PM1, PM2, UM1
TR2	TM2.1	To enable and use secure transmission protocols	PM1, UM1
	TM2.2	To include integrity solutions, such as hashing or certificates	PM1, PM2
TR3	TM3.1	To monitor all the connections established	UM1
	TM3.2	To enable and use secure transmission protocols	PM1, PM2, UM1
	TM3.3	To close unused ports	PM2, UM1
	TM3.4	To activate diode functionalities (only entrance/only release)	PM1, PM2, UM1
DR2	DM2	To include integrity solutions, such as hashing or certificates	PM1, PM2
MR1	MM1.1	To include access control features	PM1, PM2
	MM1.2	To use the principle of Need-to-know	PM2
MR2	MM2.1	To provide solid authentication control features	PM1, PM2
	MM2.2	To use hardened passwords, codes or methods of authentication to avoid impersonation based on easy-to-access private data of the user	PM1
MR4	MM4.1	To provide means of update for devices and services	PM1, PM2, UM1
MR5	MM5.1	To guarantee hardened password policies	PM1, PM2

	MM5.2	To avoid unsecure practices such as the reuse of passwords or the use of data easy to guess or to obtain	PM2, UM1
MR6	MM6.1	To guarantee S-SDLC practices	PM1, PM2
	MM6.2	To only use applications and devices provided by trusted vendors	PM1, UM2

As a consequence, there can be extracted the following conclusions:

- The methodology outcomes are aligned with the criterias and analysis of security experts, what shows that its results are reliable.

- It provides a clear context, common for any IoT environment, and an approach enough flexible to be able to be adapted to any Smart Toy use case, addressing the lack of agreement gap.

- It provides a well structured step by step procedure, that shows exactly what to consider, but also how to implement it, solving the concerns about the unclear implementation procedures.

- The language is simple and easy to understand, even by users with no deep technical knowledge. However, there is an important nuance that should be considered regarding this last point: the person implementing the methodology has to know perfectly which are the technical features and functions of the toy, to identify their risks and if they have already been mitigated by the manufacturer or not.

If that would not be the case, the subject would have two options to tackle this issue: consulting the instructions or manufacturing features of the toy (in its own packaging or on the internet, for example) until gathering enough information, or using a supporting tool that automates the analysis process.

On this behalf, a proposal would be to automate the execution of the methodology, including the elements scan, on a device with which the users can interact, such as a gateway. This device, through questions or text options, would guide the process and automatically perform the risk assessment, showing the users the resulting information and/or mitigations for the scanned Smart Toys vulnerabilities.

In this case, through a user-friendly interface, the device would manage itself the most technical-related steps and would allow the users to be informed, interact and decide about the IoT systems in a simple and intuitive manner. This would allow users to manipulate configurations and harden their Smart Toys, whenever this would be possible.

As well, in the cases that scape to the user control, it would at least inform them about the risks they may be taking in a simple language, allowing them to take in-formed decisions about their choices of using or not certain Smart Toys.

VI. CONCLUSIONS

Main Contributions

After the present research and development, there could be extracted the following conclusions:

(i) The proposed methodology addresses a current problematic existing in the IoT, and more particularly, Smart Toy industry, covering the security gaps of other existing security solutions.

(ii) It gathers all the key aspects that should be considered while identifying security risks in Smart Toys environments through an approach enough flexible to consider different types of Smart Toys, but also enough defined to be accurate on the results provided.

(iii) It also defines a security framework for a specific IoT environment that can be used as a base in further studies to develop a scalable or more detailed methodology to securitize other IoT environments.

(iv) Its flexible approach and high level of definition also provide the ability to adapt the implementation of it depending on the target audience for its use, which can be particularly interesting for two groups of people: Smart Toys manufacturers (to decide which features and elements include in their devices) and final users (to take informed decisions).

Future lines of work

Along the development of this project and as a conclusion for it, there have been identified possible and interesting future lines of work which are presented next:

(i) The possibility of complete and bring to a low-level of detail the presented development, with the purpose of providing a higher level of accuracy while identifying risks and mitigations.

(ii) To develop a commercial solution based on the proposed methodology, such as the mentioned gateway, that would allow users to easily identify the security risks that they are exposing themselves to in an IoT environment. Such a resource would not only help to prevent or mitigate active risks for children and their families, but also to spread a security information culture among the most vulnerable agents (users). Moreover, this solution could also be presented as a business solution that would allow providers of Smart Toys to identify and correct their risk sources before deploying their solutions, preventing them from great money losses as a consequence of fines or reputation damages, among others.

REFERENCES

- [1] Gartner, "Leading The IoT," 2017. [Online]. Available: https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf. [Accessed 26 February 2019]
- [2] Mozilla, "Internet Health Report 2018," 10 April 2018. [Online]. Available: <https://internethealthreport.org/2018/introduction/how-healthy-is-the-internet/>. [Accessed 26 February 2019]
- [3] ESET, "5 ways to protect your internet of things," 3 April 2017. [Online]. Available: <https://www.eset.com/us/about/newsroom/corporate-blog/survey-internet-of-stranger-things/>. [Accessed 29 February 2019]
- [4] "New rules to prevent children's 'smart' toys from being hacked," 21 November 2018. [Online]. Available: <https://www.itv.com/news/2018-11-21/new-rules-on-internet-toy-security/>. [Accessed 26 February 2019]
- [5] R. Minerva, A. Biru and D. Rotondi, "Towards a definition of the Internet of Things (IoT)," IEEE, 27 May 2015. [Online]. Available: https://iot.ieee.org/images/files/pdf/IEEE_IoT_Towards_Definition_Internet_of_Things_Revision1_27MAY15.pdf. [Accessed 26 February 2019]
- [6] ENISA, "Baseline Security Recommendations for IoT," 20 November 2017. [Online]. Available: <https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot>.
- [7] P. DeBeasi, "Architect IoT Using the Gartner Reference Model," 26 April 2018. [Online]. [Accessed June 2018].
- [8] M. Burhan, R. Asif Rehman, B. Khan and B. S. Kim, "IoT Elements, Layered Architectures and Security Issues: A Comprehensive Survey," PMC, 24 August 2018. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6165453/>. [Accessed 26 February 2019].
- [9] Federal Trade Commission, "Children's Online Privacy Protection Rule ("COPPA")," [Online]. Available: <https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule>. [Accessed 26 February 2019].
- [10] Regulation GDPR," European Parliament and the Council, 27 April 2016. [Online]. Available: <https://gdpr-info.eu/>. [Accessed 26 February 2019].
- [11] INCIBE, "INCIBE y la Asociación Española de Fabricantes de Juguetes publican una guía para el uso seguro de los juguetes conectados," INCIBE, 11 December 2018. [Online]. Available: <https://www.incibe.es/sala-prensa/notas-prensa/incibe-y-asociacion-espanola-fabricantes-juguetes-publican-guia-el-uso>. [Accessed 26 February 2019].
- [12] OWASP, "IoT Attack Surface Areas," 29 November 2015. [Online]. Available: https://www.owasp.org/index.php/IoT_Attack_Surface_Areas. [Accessed 26 February 2019].
- [13] NISTIR, "Interagency Report on Status of International Cybersecurity Standardization for the Internet of Things (IoT)," February 2018. [Online]. Available: <https://csrc.nist.gov/publications/detail/nistir/8200/draft>
- [14] R. Shahan and B. Lamos, "Internet of Things (IoT) security architecture," Microsoft Azure, 10 September 2018. [Online]. Available: <https://docs.microsoft.com/en-us/azure/iot-fundamentals/iot-security-architecture>. [Accessed 26 February 2019].
- [15] Motherboard, "One of the Largest Hacks Yet Exposes Data on Hundreds of Thousands of Kids," 27 November 2015. [Online]. Available: https://motherboard.vice.com/en_us/article/yp3z5v/one-of-the-largest-hacks-yet-exposes-data-on-hundreds-of-thousands-of-kids. [Accessed 26 February 2019]
- [16] T. Hunt, "Data from connected CloudPets teddy bears leaked and ransomed, exposing kids' voice messages," 28 February 2017. [Online]. Available: <https://www.troyhunt.com/data-from-connected-cloudpets-teddy-bears-leaked-and-ransomed-exposing-kids-voice-messages/>.
- [17] P. Stone, "Hacking Unicorns with Web Bluetooth," 28 February 2017. [Online]. Available: <https://www.contextis.com/blog/hacking-unicorns-web-bluetooth>.
- [18] Change.org, "Drop the 'Hello Barbie' Eavesdropping Doll," 2015. [Online]. Available: <https://www.change.org/p/mattel-drop-the-hello-barbie-eavesdropping-doll>. [Accessed 26 February 2019].
- [19] C. Baraniuk, "Call for privacy probes over Cayla doll and i Que toys," 6 December 2016. [Online]. Available: <https://www.bbc.com/news/technology-38222472>. [Accessed 26 February 2019]
- [20] A. Liptak, "Amazon's Alexa started ordering people dollhouses after hearing its name on TV," 7 January 2017. [Online]. Available:

<https://www.theverge.com/2017/1/7/14200210/amazon-alexa-tech-news-anchor-order-dollhouse>. [Accessed 26 February 2019].

Paula González Domínguez (paulagonzalezdom@gmail.com) is a Senior Consultant at Deloitte. She received a MSc in Cyber Security by the UNIR (Universidad Internacional de La Rioja) in 2018, with honors in her Master's Thesis. Her work and research are focused on Infrastructure Protection, Security Risk Assessment, Cybersecurity Strategy and Awareness.

Fidel Paniagua Díez (fidel.paniagua@uc3m.es) is a researcher in the Evalues Lab (IT Security Evaluation) at Carlos III University of Madrid and is pursuing a PhD in computer security. His research interests include access control models and designing, developing, and evaluating secure communication systems. Paniagua Díez's PhD work is part of the research project SAVIER (Situational Awareness Virtual Environment), supported by Airbus Defense and Space. He received a BSc in computer engineering from Carlos III University of Madrid. He is Certified Ethical Hacker and EC-Council Certified Security Analyst.

Diego Suárez Touceda (diego.suarez@unir.net) is Associate Professor and Researcher at UNIR (Universidad Internacional de La Rioja) and Key Account Manager and Sr. Cybersecurity Expert at Clover Technologies S.L. He is Ph.D. in Information Security, CISM, CISSP, CEH and ECSA. His work and research are focused on Security Architectures, Network Security Services, Access Control Systems, Cybersecurity, Wearable Devices, P2P Systems, IoT, Smart Cities and Cloud Computing.

Juan José Nombela Pérez (juanjose.nombela@unir.net) is academic director of the master's degree on Cyber Security at UNIR (Universidad Internacional de La Rioja). He received a MSc in Cyber Security from Polytechnic University of Madrid and he is CISA, CISM and 27001 Lead Auditor. His work and research are focused on Biometric Security, Network Security, Mobile Device Security, Identity Management and Security management.

Análisis de la Estructura del Contenedor de Vídeos Digitales de Dispositivos Móviles para Identificación de la Fuente de Vídeos en Escenarios Abiertos

Raquel Ramos López, Elena Almaraz Luengo, Ana Lucila Sandoval Orozco, Luis Javier García Villalba*, *Member, IEEE*

Resumen—La ciencia forense se ha servido de la tecnología multimedia para analizar, evidenciar e incluso dilucidar responsabilidades en los procesos judiciales. El análisis de vídeos digitales adquiere especial relevancia al permitir determinar tanto el origen como la autenticidad de un material y de relacionar a un individuo con un dispositivo, lugar o evento. El constante desarrollo de la tecnología hace que, a pesar de que los principios básicos sean inalterables, el análisis de vídeos digitales requiera, en el ámbito forense, de nuevos procedimientos y herramientas de enfoque. Por consiguiente, es necesario proporcionar al analista forense técnicas para identificar el contenido multimedia. En este trabajo se estudia el problema de la identificación de la fuente de vídeos en escenarios abiertos, esto es, aquéllos en los que no se conozca a priori el conjunto de cámaras a las que pertenezca el vídeo a fin de identificar su fuente, hecho éste que se produce en casos reales. En particular, se propone un algoritmo de identificación de la fuente de adquisición de vídeos digitales generados por los dispositivos móviles usando algoritmos no supervisados basados en el análisis de la estructura del contenedor de vídeo.

Palabras claves—Análisis de Conglomerados, Análisis Forense, Átomos, Contenedor de vídeo, Estructura del Contenedor, Identificación de fuente

I. INTRODUCCIÓN

El análisis forense de dispositivos móviles se ha convertido en una de las áreas de investigación más importantes. En primer lugar, las capacidades de los dispositivos inteligentes han mejorado sustancialmente, siendo más utilizados que los portátiles ya que los usuarios los tienen a su alcance en cualquier momento, registrando constantemente sus actividades y movimientos proporcionando una visión del comportamiento del usuario [1].

La combinación de teléfonos móviles inteligentes con plataformas de social media y almacenamiento en la nube ha permitido que el vídeo se convierta en una importante fuente de información. Estos vídeos digitales se pueden realizar en cualquier momento y lugar para diferentes propósitos y distribuir en Internet en un corto período de tiempo. En ocasiones su contenido puede estar relacionado con actos ilegales como terrorismo, pornografía infantil, espionaje industrial, etc.

R. Ramos López Securitas Direct (rlopez@securitasdirect.es). R. Ramos López, E. Almaraz Luengo, A. L. Sandoval Orozco and L. J. García Villalba. Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Ingeniería del Software e Inteligencia Artificial, Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, España. {raqram01, ealmaraz}@ucm.es, {asandoval, javiergv}@fdi.ucm.es.

Para abordar estos problemas, los investigadores han desarrollado algoritmos forenses que verifican la autenticidad y la fuente del contenido digital [2]. Las técnicas forenses que identifican información sobre la fuente cuando se genera contenido multimedia (imágenes o vídeos), se dividen en dos grupos: por un lado, aquéllas orientadas a la verificación del origen de un contenido multimedia y aquéllas destinadas a la detección de inconsistencias en la fuente dentro del contenido multimedia [2], [3].

Existen diversas investigaciones que desarrollan algoritmos forenses para determinar la identificación de la fuente de una imagen aunque los estudios son escasos en el caso de los vídeos digitales [3]. En [4] se sugiere que estos algoritmos utilizan trazas dejadas por una amplia variedad de componentes físicos y algorítmicos en la tubería de procesamiento de una cámara. Se han diseñado algoritmos de modelos de cámaras forenses que aprovechan las huellas dejadas por la demosaicing (método de los artefactos CFA y Demosaicing), [5], [6], [7], [8], [9] e información del encabezado JPEG [10]. La mayor parte del trabajo existente se ha centrado en el uso de huellas dactilares de sensores para identificar el dispositivo fuente específico de un vídeo, por ejemplo, ver [11], [12], [13], [14], [15] [16].

El análisis de la fuente de adquisición de vídeo es uno de los primeros problemas que han surgido en las técnicas de análisis forense. Dentro de la identificación de la fuente de adquisición existen dos enfoques principales: escenarios cerrados o escenarios abiertos. Un escenario cerrado es aquél en el que la identificación de la fuente del vídeo se realiza en un conjunto de cámaras específicas y conocidas. Para este enfoque, normalmente se utiliza un conjunto de vídeos de cada dispositivo para formar a un clasificador y, posteriormente, se predice la fuente de adquisición de los vídeos que se están investigando.

En [17] se presenta un esquema de identificación de fuentes de vídeo digital basado en máquinas de soporte vectorial (SVM) y ruido PRNU. Con un vídeo de entrada, los fotogramas con cambios de escena más significativos se extraen utilizando el histograma de color. Un total de 81 funciones, que son los componentes Wavelet del sensor, se utilizan para entrenar al clasificador SVM con vídeos de entrenamiento. Un total de 5 dispositivos diferentes de 5 marcas diferentes fueron utilizados para entrenar al clasificador SVM. Los resultados obtenidos muestran una tasa de éxito del 87 % o del 90 %, dependiendo de la resolución del vídeo.

En un escenario abierto, no se conoce inicialmente el conjunto de dispositivos a los que pertenecen los vídeos para identificar su fuente de adquisición. El objetivo no es identificar la marca y el modelo de los vídeos, sino poder agruparlos en grupos en los que todos sus vídeos pertenezcan al mismo dispositivo. Este último enfoque es más realista, ya que en muchos casos el analista ignora completamente el conjunto de dispositivos a los que puede pertenecer un conjunto de vídeos. Identificar el dispositivo que genera el contenido digital es muy importante en el contexto de un proceso judicial porque puede incriminar o delimitar responsabilidades a un sospechoso antes de un acto delictivo. Para realizar cualquier tipo de clasificación de vídeo en escenarios abiertos o cerrados, es necesario obtener ciertas características que permitan a las técnicas de clasificación realizar su tarea. Dentro del análisis forense de vídeo digital, las principales técnicas de análisis se dividen en 5 grupos: 1) Metadatos, 2) Características de la imagen, 3) Defectos de la matriz CFA, 4) Imperfecciones del sensor y 5) Características de los contenedores multimedia.

Este trabajo propone una técnica de identificación de la fuente de adquisición de vídeos digitales generados por dispositivos móviles mediante el uso de algoritmos no supervisados basados en el análisis de la estructura de los contenedores de vídeo multimedia. Se ha dividido en 5 secciones, la primera consistente en esta introducción. En la Sección II se presentan las principales técnicas actuales de análisis en este contexto. El algoritmo propuesto en este trabajo se desarrolla en la Sección III. En la Sección IV se muestran los resultados numéricos en los que se ha aplicado el algoritmo diseñado. Finalmente, en la Sección V se presentan las conclusiones extraídas de este trabajo.

II. HERRAMIENTAS DE ANÁLISIS FORENSE DE VÍDEO

Un vídeo está formado por una secuencia de imágenes llamadas fotogramas que varían con el tiempo dando una sensación de movimiento. Debido al gran volumen de información que tiene un vídeo, éste se codifica y decodifica mediante un algoritmo matemático conocido como códec. A su vez, estas tramas ya codificadas se encapsulan junto con las pistas de audio, metadatos y subtítulos en un único archivo conocido como contenedor multimedia. En la Tabla I se muestran los diferentes elementos que componen un contenedor multimedia.

Tabla I
ELEMENTOS COMPOSITIVOS DE UN CONTENEDOR MULTIMEDIA.

Formato del contenedor: .avi, .mp4, .mov, .ogg, .flv, .mkv, etc			
Vídeo códec	Audio códec	Captioning descripción de vídeo	Metadatos
H.264.	AAC	SAMI	MPEG-7
VC-1	WMA	SMIL	CableLabs
Theora	Vorbis	Hi-Caption	TV-Anytime
Dirac 2.1	PCM, etc.	CMML	EBU
H.263, etc.		DFXP	XPM, etc.
		3GPP TS	
		MPSUB, etc.	

Los contenedores multimedia o formatos de vídeo se definen como aplicaciones informáticas capaces de almacenar

audio y vídeo y, en algunos casos, también subtítulos y otra información adicional.

Los contenedores multimedia más utilizados en la actualidad son:

- AVI (Audio Video Interleave): contenedor multimedia estándar de Windows.
- MP4: contenedor estandarizado para MPEG4 [18].
- FLV (Flash Video): formato utilizado para entregar vídeo MPEG a través de Flash Player.
- MKV (Mastroska): contenedor de especificaciones abierto orientado a animación.
- MOV: formato de contenedor QuickTime de Apple.
- OGG, OGM, OGV: contenedores estándar abiertos.

En la literatura más reciente, se puede encontrar que la mayoría de las investigaciones analizan la estructura interna de los contenedores multimedia en el caso del formato AVI, siendo casi inexistente el estudio de los contenedores MP4, 3GP y MOV.

Uno de los primeros trabajos donde se realiza un análisis detallado de las estructuras de los vídeos es [19] donde se analizan en detalle las secuencias de vídeo AVI y MP4 (MOV, 3GP, MP4) de teléfonos móviles y cámaras digitales. Uno de los principales resultados que se obtienen es que los vídeos de cámaras digitales y teléfonos móviles suelen emplear diferentes formatos de contenedores y codecs de compresión. Los teléfonos móviles optan por sofisticados algoritmos de compresión (MP4V, H.26x). La mayoría de las cámaras digitales de nuestro equipo de prueba prefieren una combinación de contenedores AVI y compresión básica MPEG. La estructura de los contenedores tipo AVI y MP4 no está estrictamente definida. Observaron diferencias considerables tanto en el orden como en la presencia de segmentos de datos individuales. Los archivos AVI a menudo contienen listas de información específicas o trozos de JUNK. Los archivos de tipo MP4 pueden emplear varios átomos no estándar y diferentes parametrizaciones de entradas de átomos específicos. La edición de vídeo sin pérdida deja intactos los ajustes de compresión del flujo de vídeo original, pero introduce sus propios artefactos distintivos en la estructura de los archivos de contenedor. Mientras que las peculiaridades del formato de archivo del dispositivo fuente original se pierden normalmente después de la edición de vídeo, todas las herramientas de software probadas tienen firmas de formato de archivo únicas en todo su conjunto de pruebas.

En [20] se introduce un método para el análisis no supervisado de contenedores de archivos de vídeo y sus autores presentan dos aplicaciones forenses principales de dicho método: la verificación de la integridad del vídeo (basado en la diferencia entre un contenedor de archivos de referencia y uno de consulta) y la identificación y clasificación de la marca del dispositivo de origen (basado en el análisis de la estructura y contenido de los contenedores). Se comprobó la eficacia de ambas aplicaciones en un conjunto de datos compuesto por 578 vídeos tomados con smartphones modernos de las principales marcas y modelos y se llegó a la conclusión de que la solución propuesta ofrece un coste computacional extremadamente bajo en comparación con todas las técnicas disponibles basadas en el análisis del flujo de vídeo o la inspección manual de los contenedores de archivos.

En [21] se investiga el contenido de vídeo almacenado en Video Event Data Recorders (VEDRs). En concreto, se estudia la estructura de los archivos de vídeo para cada tipo de software de edición de vídeo que dejaría huellas del procesamiento del software de edición de vídeo. Debido a que tales trazas son una característica inherente a cada paquete de software de edición de vídeo, pueden detectar el software de edición de vídeo específico utilizado para manipular el vídeo, además de si el vídeo fue, de hecho, manipulado. Para evaluar la precisión de su técnica, examinaron 296 archivos de vídeo no modificados de Audio Video Interleave (AVI). Se realizó este examen utilizando versiones populares de software de edición de vídeo. Como resultado, se encontró que las estructuras de datos AVI en los archivos de vídeo modificados aparecen consistentemente de acuerdo con cada paquete de software de edición de vídeo. Cada estructura de datos resultante no se ve afectada por la estructura del archivo de vídeo original.

III. DESCRIPCIÓN DEL ALGORITMO PROPUESTO

Este artículo presenta una técnica para identificar la fuente de adquisición de vídeo digital generada por los dispositivos móviles. La técnica se divide en dos tareas: 1) Extracción de la información de los átomos contenidos en cada uno de los vídeos y 2) Agrupación mediante técnicas de análisis de conglomerados de cara a la identificación del vídeo por modelo.

Para la realización de las tareas se han tenido en cuenta los formatos de contenedores MOV basados en los estándares QuickTime de Apple QuickTime [22], MP4 y 3GP compatibles con el estándar ISO / IEC 14496 Parte 12 [18]. Se han analizado los contenedores que existen tanto en la plataforma principal de compartición de vídeo, YouTube, como en una de las principales plataformas de mensajería de WhatsApps porque se encuentran dentro de uno de los conjuntos de datos utilizados para este trabajo [23].

III-A. Análisis de la Estructura del Contenedor

La estructura elemental de un vídeo es el átomo. Los metadatos, el vídeo y el sonido de un vídeo están dentro de ellos. Los átomos son de naturaleza jerárquica. Es decir, un átomo puede contener otros átomos, que pueden contener otros y así sucesivamente. El tipo de átomo se especifica mediante un número entero sin signo de 32 bits, normalmente interpretado como un código ASCII (American Standard Code for Information Interchange) de cuatro caracteres, generalmente en minúsculas. Debe tenerse en cuenta que no hay ninguna regla respecto a los átomos que deben aparecer y su orden, sin embargo, la mayoría sigue una estructura similar [24]. Este algoritmo se ha utilizado para extraer información múltiple de cualquier formato de vídeo como: Vídeo MP4 / H.264, MOV y 3GP, para extraer información de los átomos.

La extracción de átomos consiste en almacenar las etiquetas, valores, orden de aparición de los átomos y todo tipo de información relevante de un vídeo digital generado por un dispositivo móvil. El proceso comienza con la obtención del byte inicial del átomo, tamaño y tipo de átomo con una longitud máxima de 4 bytes formados como una cadena de

caracteres (ejemplo: ftyp, moov, mdat, etc.). A continuación se verifica la duplicidad de los átomos y la existencia de átomos hijos. Finalmente, se obtiene un diccionario de un conjunto de átomos y etiquetas (Path-tag) con sus respectivos valores y órdenes de apariencia. Para un estudio más profundo de los átomos, ver [19] y [24].

La Tabla II muestra el output que se obtiene tras la extracción de los átomos. El primer átomo es “ftyp” tal y como se indica en las especificaciones de [19]. Como los átomos se organizan jerárquicamente (*ie./moov/*), a su vez tienen átomos hijos (*ie./moov/trak*) y etiquetas (*ie./moov/mvhd/tkhd/flags*) que a su vez contienen valores (*ie./moov/mvhd/tkhd/version,value:0*).

Tabla II
MUESTRA DE LA INFORMACIÓN EXTRAÍDA DEL CONTENEDOR

Path	PathOrder	Field	Value
ftyp	ftyp-1	majorBrand	mp42
ftyp	ftyp-1	minorVersion	1
ftyp	ftyp-1	compatibleBrands	mp41mp42isom
beam	beam-2	byteInitial	28
beam	beam-2	size	42
moov/trak/tkhd	moov-4/trak-2/tkhd-1	version	0
moov/trak/tkhd	moov-4/trak-2/tkhd-1	flags	1
moov/trak/tkhd	moov-4/trak-2/tkhd-1	trackId	1
moov/trak/tkhd	moov-4/trak-2/tkhd-1	trackWidth	48
moov/trak/tkhd	moov-4/trak-2/tkhd-1	trackHeight	848

En esta propuesta, para realizar la agrupación de vídeos se ha tenido en cuenta que los vídeos son conjuntos de elementos que contienen las siguientes características:

- **PathField:** Se define como la unión de las etiquetas Path y Field separadas por el carácter ‘/’. En la Tabla II, el campo PathField de la primera fila corresponde al valor (*ftyp/majorBrand*).
- **PathFieldValue:** Se define como la unión de las etiquetas Path y Field separadas por ‘/’ y se añade la etiqueta Value separada por ‘=’. En la Tabla II, el campo PathFieldValue de la primera fila corresponde a: *ftyp/majorBrand = mp42*.
- **PathOrderField:** Se define como la unión de las etiquetas PathOrder y Field separadas por ‘/’. En la Tabla II, el campo PathOrderField de la primera fila corresponde con: *ftyp-1/majorBrand*.
- **PathOrderFieldValue:** Es la unión de las etiquetas PathOrder y Field separadas por ‘/’ al que se le añade la etiqueta Value separada por ‘=’. En la Tabla II, el campo PathOrderFieldValue de la primera fila se corresponde con: *ftyp-1/majorBrand = mp42*.

III-B. Técnicas de Análisis de Conglomerados

El análisis de conglomerados es un conjunto de técnicas para agrupar las observaciones por afinidad. Esta es la razón por la que esta técnica ha sido tradicionalmente considerada como parte de la Estadística Multivariante, aunque actualmente tiende a ser catalogada como Minería de Datos. En estas técnicas no se hacen suposiciones sobre el número de grupos o la estructura del grupo. La agrupación se hace sobre la base de similitudes o distancias (diferencias).

Cuando los artículos (unidades, cajas) están agrupados, la proximidad es usualmente indicada por algún tipo de distancia.

Definición 1: La distancia $d(P, Q)$ entre dos puntos P y Q satisface las siguientes propiedades, donde R es otro punto intermedio:

1. $d(P, Q) = d(Q, P)$.
2. $d(P, Q) > 0$; si $P \neq Q$.
3. $d(P, Q) = 0$; si $P = Q$.
4. $d(P, Q) \leq d(P, R) + d(R, Q)$.

La tercera condición de la definición 1 se conoce como desigualdad triangular. Aquellas medidas que verifican las condiciones 1 y 2 de la definición 1 pero no la tercera, se conocen como semimétricas. Existen muchos tipos de distancia que se utilizan en el análisis de conglomerados [25]. Sean $\mathbf{x}' = (x_1, \dots, x_p)$ y $\mathbf{y}' = (y_1, \dots, y_p)$.

- Distancia Euclídea: $d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})}$.
- Distancia Estadística: $d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})' \mathbf{A} (\mathbf{x} - \mathbf{y})}$. Generalmente, $\mathbf{A} = \mathbf{S}^{-1}$, donde $\mathbf{S}$ contiene las varianzas y covarianzas muestrales. Sin embargo, sin tener un conocimiento previo de los diferentes grupos, esas cantidades muestrales no se pueden computar, es por ello que en el análisis de conglomerados se prefiere usar la distancia Euclídea.
- Métrica de Minkowski: $d(\mathbf{x}, \mathbf{y}) = (\sum_{i=1}^p |x_i - y_i|^m)^{1/m}$. Para $m = 1$, $d(\mathbf{x}, \mathbf{y})$ mide la distancia en manzanas de ciudad (city-block) entre dos puntos en dimensión p . Para $m = 2$, $d(\mathbf{x}, \mathbf{y})$ es la distancia Euclídea.
- Otras: Métrica de Canberra, coeficiente de Czekanowski (para el caso de variables no negativas).

Por otro lado, cuando los ítems no se pueden representar de manera adecuada por una medida p -dimensional, cada par de ítem se comparan en base a si tienen o no determinada característica. En este caso, se utilizan variables binarias en las que 1 indicará la presencia de la característica y 0 su ausencia. Esta situación se representa mediante una tabla de contingencia.

Las técnicas de análisis de conglomerados se pueden clasificar en dos clases: métodos jerárquicos y no jerárquicos. En este artículo se usarán técnicas de clasificación jerárquicas.

Los métodos jerárquicos proceden, bien de una serie de fusiones sucesivas, bien de una serie de escisiones sucesivas:

- Métodos jerárquicos aglomerativos: parten de los elementos individuales y los añaden en grupos.
- Métodos jerárquicos divisivos: parten de un conjunto formado por todos los elementos y lo van dividiendo sucesivamente hasta llegar a alcanzar los elementos individuales.

Los algoritmos aglomerativos tienen siempre la misma estructura y solamente difieren en la manera en que calculan la distancia entre grupos. Su estructura se muestra en el Algoritmo 1.

Algoritmo 1: Algoritmo aglomerativo

- ① Comenzar con N conglomerados, cada uno conteniendo un ítem y una matriz de dimensión $N \times N$ de distancias (o similitudes) entre ítems $\mathbf{D} = (d_{ik})$;
 - ② Seleccionar los dos elementos más cercanos según las distancias actuales en $\mathbf{D}$ y formar con ellos una nueva clase;
 - ③ Sustituir los dos elementos utilizados en (2) para definir la clase por un nuevo elemento que represente la clase construida. Las distancias entre este nuevo elemento y el elemento se calculan utilizando uno de los criterios que se exponen a continuación;
 - ④ Volver al paso (2) y repetir (2) y (3) hasta conseguir una única clase;
-

Como se mencionó anteriormente, existen diferentes criterios para medir distancias. Los más comunes se pueden ver en la Tabla III en la que $d(u, v)$ representa la distancia entre los elementos u y v , d_{UW} y d_{VW} son las distancias entre los vecinos más próximos de los conglomerados U y W y V y W respectivamente y $|U|$ y $|V|$ representan el cardinal de U y V respectivamente.

El resultado de los métodos aglomerativos y divisivos se puede representar en un diagrama bidimensional conocido como dendograma que muestra las sucesivas uniones (o divisiones) que se han realizado en cada iteración del algoritmo. Una vez que se obtiene el dendograma, se pueden obtener los diferentes conglomerados.

Otra metodología de análisis de conglomerados en el espacio es la basada en la densidad de aplicaciones con ruido, conocida usualmente con sus siglas en inglés DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Esta metodología fue propuesta por Martin Ester, Hans-Peter Kriegel, Jörg Sander y Xiaowei Xu en 1996 [26].

Considérese un conjunto de puntos que se desea agrupar en un determinado espacio. El algoritmo DBSCAN clasifica los puntos en tres tipos: puntos núcleo, puntos densamente-alcanzables o ruido.

Definición 2: (Puntos núcleo, directamente alcanzable y ruido)

- Un punto p es un punto núcleo si al menos $minPoints$ puntos están a una distancia ϵ de él (incluyendo p), y esos puntos son directamente alcanzables desde él. No es posible tener puntos directamente alcanzables desde un punto que no sea un núcleo.

Tabla III
CRITERIOS MÁS COMUNES DE VINCULACIÓN EN ANÁLISIS JERÁRQUICO DE CONGLOMERADOS

Criterio	Descripción: la distancia entre dos grupos nuevos es la	Fórmula
Encadenamiento simple	menor de las distancias entre grupos antes de la fusión	$d_{(UV),W} = \min \{d_{UW}, d_{VW}\}$
Encadenamiento completo	mayor de las distancias entre los grupos antes de la fusión	$d_{(UV),W} = \max \{d_{UW}, d_{VW}\}$
Media ponderada	semisuma de las distancias entre grupos antes de la fusión	$d((U \cup V), W) = \frac{d(U,W) + d(V,W)}{2}$
Promedio no ponderado	dist. media entre todos los pares (i, j) con i en un grupo y j en otro	$\frac{1}{ U V } \sum_{u \in U} \sum_{v \in V} d(u, v)$

- Un punto q es directamente alcanzable desde p si existe una secuencia de puntos $p_1, \dots, p_n$ donde $p_1 = p$ y $p_n = q$, tales que, cada punto p_{i+1} es directamente alcanzable desde p_i .
- Un punto que no sea alcanzable desde cualquier otro punto es considerado ruido.

Si p es un punto núcleo, éste forma un conglomerado con otros puntos (núcleo o no) que sean alcanzables desde él.

Cada conglomerado contiene al menos un punto núcleo. Los puntos no núcleos alcanzables pueden pertenecer a un conglomerado pero actúan como una barrera puesto que no es posible alcanzar más puntos desde estos. Obsérvese que la propiedad de ser alcanzable no es simétrica.

Por definición, ningún punto puede ser alcanzable desde un punto que no sea núcleo, independientemente de la distancia a la que se encuentre, por tanto es necesario definir el concepto de conectividad para especificar la noción de conglomerado usada en el algoritmo DBSCAN.

Definición 3: Dos puntos p y q están conectados densamente si existe otro punto o tal que p y q son directamente alcanzables desde o . Esta propiedad es simétrica.

El algoritmo DBSCAN utiliza dos parámetros: ϵ (que especifica cuán cerca deben estar los puntos entre sí para ser considerados parte de un mismo conglomerado) y $minPoints$ (que representa el número mínimo de puntos para formar una región densa). Los pasos fundamentales de esta metodología se muestran en el Algoritmo 2.

Algoritmo 2: Algoritmo DBSCAN

- ① Para cada punto p_i calcular la distancia $d(p_i, p_j)$, $\forall j \neq i$. Encontrar todos los puntos vecinos en un radio de ϵ del punto de partida p_i . Cada punto, con un vecino cuya distancia sea mayor o igual que $minPoints$, está marcado como punto núcleo o punto visitado;
 - ② Para cada punto núcleo que aún no haya sido asignado a un conglomerado, crear un nuevo conglomerado. Encontrar de manera iterativa todos los puntos conectados densamente y asignarlos al mismo conglomerado que el punto núcleo;
 - ③ Iterar a través de los puntos no visitados restantes en el conjunto de datos;
-

El algoritmo OPTICS (Ordering Points to Identify the Clustering Structure) se puede considerar como una generalización del algoritmo DBSCAN en el caso de múltiples rangos, reemplazando el parámetro ϵ por el radio máximo de búsqueda. Fue presentado por Michael Ankerst, Markus M. Breunig, Hans-Peter Kriegel y Jörg Sander en 1999 [27]. La idea principal es similar a la de DBSCAN pero aborda el problema de detectar conglomerados significativos en datos de densidad variable, en concreto, OPTICS calcula un orden de los puntos aumentados por información adicional, es decir, la distancia de alcanzabilidad, que representa la estructura jerárquica intrínseca del conglomerado. A continuación mostraremos la propuesta específica de este trabajo.

- Extracción y almacenamiento de datos: para extraer la información de los átomos se utilizará una solución escrita en Python y desarrollada por el Grupo de Análi-

sis, Seguridad y Sistemas (GASS) del departamento de Ingeniería del Software e Inteligencia Artificial de la Universidad Complutense de Madrid. Esta solución es capaz de analizar información múltiple de cualquier vídeo MP4/H.264, MOV y 3gp, para extraer la información de los átomos. Los valores del interior han sido convertidos a utf-8 cuando ha sido posible y de lo contrario permanecen como hexadecimales.

- Recopilar los datos: con la información obtenida, consultar la base de datos para descargar los vídeos necesarios. Posterior agrupación de los mismos para obtener un único cuadro de datos.
- Vectorización: agrupación de los datos con el objetivo de que el uso posterior de los mismos sea más abarcable y facilite la manipulación de los mismos.
- Filtrado: mediante pandas, filtrar las columnas correspondientes a los átomos y campos no deseados. Este filtrado es una extensión de la realizada en [20].
- Procedimiento de Análisis de Conglomerados: recuperar la matriz de valores binarios del marco de datos que representa el conjunto de datos. Cada línea contiene una observación, un archivo de vídeo y dimensiones de columna. Para agrupar los datos, se proponen dos algoritmos de clasificación jerárquica y OPTICS. Los detalles se pueden consultar en el Algoritmo 3.

Algoritmo 3: Algoritmo propuesto

- ① Extracción y almacenamiento de los datos;
 - ② Recopilación los datos;
 - ③ Vectorización;
 - ④ Filtrado;
 - ⑤ Agrupación: algoritmos de conglomerados jerárquicos y OPTICS;
-

IV. EXPERIMENTOS Y RESULTADOS

IV-A. Conjuntos de Datos

Para llevar a cabo los experimentos se han utilizado los dos conjuntos de datos más recientes en la literatura, concretamente: VISION dataset [23] y ACID dataset [4]. En concreto se han utilizado muestras de ambos conjuntos de datos. La información concreta de cada muestra se puede ver en las Tablas IV y V.

IV-B. Condiciones Experimentales

Como se comentó anteriormente para el procedimiento de obtención de conglomerados, es necesario definir una medida. Sin embargo, en un procedimiento de análisis de conglomerados, es posible incluso trabajar con semimétricas. En particular, las medidas utilizadas en nuestro trabajo han sido: Euclídea, correlación, Rogers-Tanimoto y métrica de Sokal-Sneath.

Para la realización de los experimentos se han tenido en cuenta ciertas consideraciones. En primer lugar, la etiqueta Field no siempre es válida para identificar el origen, ya que tiene valores específicos que dependen del propio vídeo, en el caso de las etiquetas relacionadas con la fecha de creación,

Tabla IV
COMPOSICIÓN DE LA MUESTRA DE ACID

Marca	Modelo	Dispositivo	# Vídeos
Apple	iPhone 8 plus	M00	223
Asus	Zenfoe 4 Laser	M01	239
Canon	VIXIA HF R800	M06	25
Google	Pixel 1	M10	25
	Pixel 2	M11	25
Huawei	Honor 6X Pixel 2	M12	25
	Honor Mate SE 2	M13	25
Kodak	Ektra	M15	25
	Q6	M16	25
LG	X Charge	M17	25
Moto	E4	M18	25
	G5 plus	M19	25
Nikon	Coolpix S33	M20	25
	Coolpix S3700	M21	25
	Coolpix S7000	M22	25
Olympus	Stylus Tough TG-860	M24	25
Samsung	Galaxy J7 Pro	M27	25
	Galaxy S5	M29	25
	Galaxy S7	M30	25
	Galaxy Tab A	M31	25

Tabla V
COMPOSICIÓN DE LA MUESTRA DE VISION

Marca	Modelo	# Vídeos
Apple	iPad2	16
	Ipad mini	16
	iPhone 4	19
	iPhone 4S	28
Asus	Zenphone 2 Laser	19
Huawei	Ascend G6-U10	19
	Honor 5C NEM-L51	19
	P8 GRA-L09	19
	P9 EVA-L09	19
	P9 Lite VNS-L31	19
Lenovo	Lenovo P70-A	19
LG	D290	19
Microsoft	Lumia 640 LTE	10
OnePlus	A3000	19
	A3003	19
Samsung	Galaxy S III Mini GT-I8190	16
	Galaxy S III Mini GT-I8190N	22
	Galaxy S3 GT-I9300	19
	Galaxy S4 mini GT-I9195	19
	Galaxy S5 SM-G900F	19
	Galaxy Tab 3 GT-P5210	37
	Galaxy Tab A SM-T555	16
	Galaxy Trend Plus GT-S7580	16
Sony	Xperia Z1 Compact D5503	19
Wiko	Ridge 4G	11
WhatsApp	WhatsApp	644
Xiaomi	Redmi Note 3	19
Youtube	Youtube	622

la duración, etc. Se han eliminado algunos átomos: *modificationTime*, *creationTime*, *entryCount*, *sampleCount*, *freeSpace* y *duration*. En segundo lugar, como universo se han definido todas las representaciones posibles de las etiquetas de los átomos, específicamente las siguientes: PathField, PathFieldValue, PathOrderField y PathOrderFieldValue.

Para elegir la mejor representación del conjunto de datos a agrupar y la mejor métrica, las alternativas se han evaluado siguiendo el criterio del Coeficiente de Silueta. La represen-

tación y medida con el mayor coeficiente de silueta será la más probable de ser correctamente separada.

Definición 4: El Coeficiente de Silueta es una medida de la consistencia de los conglomerados. Mide tanto la cohesión como la separación de los mismos. Sean C_i , $i = 1, \dots, k$ los conglomerados. Dado $i \in C_i$, sean

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, j \neq i} d(i, j) \quad (1)$$

la distancia media entre i y todos los demás puntos de datos en el mismo conglomerado, donde $d(i, j)$ es la distancia entre i y j en el conglomerado C_i , y

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j) \quad (2)$$

la distancia media más pequeña de i a todos los puntos de cualquier otro grupo, del cual i no es miembro. El coeficiente de silueta se define como:

$$s(i) = \begin{cases} 1 - a(i)/b(i), & \text{si } a(i) < b(i) \\ 0, & \text{si } a(i) = b(i) \\ a(i)/b(i) - 1, & \text{si } a(i) > b(i) \end{cases} \quad (3)$$

El Coeficiente de Silueta se ha utilizado ampliamente en otros trabajos de análisis forense multimedia como por ejemplo [28], [29] o [30].

En la Tabla VI se muestra el Coeficiente de Silueta máximo (para cualquier métrica) en cada uno de los conjuntos de datos utilizados en la experimentación. Además, las Tablas VII y VIII muestran el resultado de las 4 mejores métricas que han dado buenos resultados en ambos conjuntos de datos que han sido calculados usando el Coeficiente de Silueta.

Tabla VI
COEFICIENTE DE SILUETA MEDIO MÁXIMO PARA CADA MÉTRICA EN LOS CONJUNTOS DE DATOS DE LA EXPERIMENTACIÓN

Universo	Clase	Coeficiente de Silueta	
		VISION	ACID
PathField	Marca	0.042925	0.532204
PathField	Dispositivo	-0.044301	0.405100
PathField	Modelo	-0.104576	0.405133
PathFieldValue	Marca	0.490063	0.586225
PathFieldValue	Dispositivo	0.492146	0.653571
PathFieldValue	Modelo	0.465963	0.653581
PathOrderField	Marca	0.747372	0.907504
PathOrderField	Dispositivo	0.666355	0.805020
PathOrderField	Modelo	0.609535	0.805028
PathOrderFieldValue	Marca	0.585747	0.814979
PathOrderFieldValue	Dispositivo	0.538426	0.795484
PathOrderFieldValue	Modelo	0.485761	0.795490

IV-C. Evaluación del Desempeño del procedimiento de Análisis de Conglomerados

Existen numerosas medidas de comparación de los resultados de un procedimiento de análisis de conglomerados [31]. En este trabajo se hará uso del Índice Rand (RI) o medida de Rand. Este índice es un valor perteneciente al intervalo [0, 1] que calcula una medida de similitud entre dos conglomerados considerando todos los pares de muestras y contando los pares que se asignan en los mismos o diferentes conglomerados en los conglomerados predichos y verdaderos.

Tabla VII
TOP MÉTRICAS MUESTRA DE VISION

Universo	Clase	Métrica	Resultado
PathField	Marca	Euclídea	0.0429
PathField	Marca	Correlación	-0.0029
PathField	Marca	Rogers-Tanimoto	0.0047
PathField	Marca	Sokal-Sneath	0.0103
PathField	Dispositivo	Euclídea	-0.0443
PathField	Dispositivo	Correlación	-0.0504
PathField	Dispositivo	Rogers-Tanimoto	-0.0503
PathField	Dispositivo	Sokal-Sneath	-0.0488
PathField	Modelo	Euclídea	-0.1045
PathField	Modelo	Correlación	-0.1123
PathField	Modelo	Rogers-Tanimoto	-0.1127
PathField	Modelo	Sokal-Sneath	-0.1112
PathFieldValue	Marca	Euclídea	0.3073
PathFieldValue	Marca	Correlación	0.4900
PathFieldValue	Marca	Rogers-Tanimoto	0.4782
PathFieldValue	Marca	Sokal-Sneath	0.4021
PathFieldValue	Dispositivo	Euclídea	0.3042
PathFieldValue	Dispositivo	correlation	0.4921
PathFieldValue	Dispositivo	Rogers-Tanimoto	0.4860
PathFieldValue	Dispositivo	Sokal-Sneath	0.4137
PathFieldValue	Modelo	Euclídea	0.2919
PathFieldValue	Modelo	Correlación	0.4659
PathFieldValue	Modelo	Rogers-Tanimoto	0.4586
PathFieldValue	Modelo	Sokal-Sneath	0.3940
PathOrderField	Marca	Euclídea	0.7473
PathOrderField	Marca	Correlación	0.7253
PathOrderField	Marca	Rogers-Tanimoto	0.7211
PathOrderField	Marca	Sokal-Sneath	0.7329
PathOrderField	Dispositivo	Euclídea	0.6663
PathOrderField	Dispositivo	Correlación	0.6267
PathOrderField	Dispositivo	Rogers-Tanimoto	0.6337
PathOrderField	Dispositivo	Sokal-Sneath	0.6480
PathOrderField	Modelo	Euclídea	0.6095
PathOrderField	Modelo	Correlación	0.5647
PathOrderField	Modelo	Rogers-Tanimoto	0.5705
PathOrderField	Modelo	Sokal-Sneath	0.5849
PathOrderFieldValue	Marca	Euclídea	0.3838
PathOrderFieldValue	Marca	Correlación	0.5857
PathOrderFieldValue	Marca	Rogers-Tanimoto	0.5711
PathOrderFieldValue	Marca	Sokal-Sneath	0.4820
PathOrderFieldValue	Dispositivo	Euclídea	0.3545
PathOrderFieldValue	Dispositivo	Correlación	0.5379
PathOrderFieldValue	Dispositivo	Rogers-Tanimoto	0.5384
PathOrderFieldValue	Dispositivo	Sokal-Sneath	0.4673
PathOrderFieldValue	Modelo	Euclídea	0.3230
PathOrderFieldValue	Modelo	Correlación	0.4857
PathOrderFieldValue	Modelo	Rogers-Tanimoto	0.4823
PathOrderFieldValue	Modelo	Sokal-Sneath	0.4239

Tabla VIII
TOP MÉTRICAS MUESTRA DE ACID

Universo	Clase	Métrica	Resultado
PathField	Marca	Euclídea	0.5322
PathField	Marca	Correlación	0.5238
PathField	Marca	Rogers-Tanimoto	0.5238
PathField	Marca	Sokal-Sneath	0.5243
PathField	Dispositivo	Euclídea	0.4046
PathField	Dispositivo	Correlación	0.4048
PathField	Dispositivo	Rogers-Tanimoto	0.4050
PathField	Dispositivo	Sokal-Sneath	0.4050
PathField	Modelo	Euclídea	0.4051
PathField	Modelo	Correlación	0.4048
PathField	Modelo	Rogers-Tanimoto	0.4050
PathField	Modelo	Sokal-Sneath	0.4050
PathFieldValue	Marca	Euclídea	0.4127
PathFieldValue	Marca	Correlación	0.5862
PathFieldValue	Marca	Rogers-Tanimoto	0.5781
PathFieldValue	Marca	Sokal-Sneath	0.5349
PathFieldValue	Dispositivo	Euclídea	0.4521
PathFieldValue	Dispositivo	Correlación	0.6535
PathFieldValue	Dispositivo	Rogers-Tanimoto	0.6448
PathFieldValue	Dispositivo	Sokal-Sneath	0.5999
PathFieldValue	Modelo	Euclídea	0.4521
PathFieldValue	Modelo	Correlación	0.6535
PathFieldValue	Modelo	Rogers-Tanimoto	0.6448
PathFieldValue	Modelo	Sokal-Sneath	0.5999
PathOrderField	Marca	Euclídea	0.8945
PathOrderField	Marca	Correlación	0.8921
PathOrderField	Marca	Rogers-Tanimoto	0.8934
PathOrderField	Marca	Sokal-Sneath	0.9075
PathOrderField	Dispositivo	Euclídea	0.8045
PathOrderField	Dispositivo	Correlación	0.8050
PathOrderField	Dispositivo	Rogers-Tanimoto	0.8049
PathOrderField	Dispositivo	Sokal-Sneath	0.8047
PathOrderField	Modelo	Euclídea	0.8045
PathOrderField	Modelo	Correlación	0.8050
PathOrderField	Modelo	Rogers-Tanimoto	0.8049
PathOrderField	Modelo	Sokal-Sneath	0.8047
PathOrderFieldValue	Marca	Euclídea	0.6659
PathOrderFieldValue	Marca	Correlación	0.8205
PathOrderFieldValue	Marca	Rogers-Tanimoto	0.8149
PathOrderFieldValue	Marca	Sokal-Sneath	0.7622
PathOrderFieldValue	Dispositivo	Euclídea	0.6196
PathOrderFieldValue	Dispositivo	Correlación	0.7954
PathOrderFieldValue	Dispositivo	Rogers-Tanimoto	0.7896
PathOrderFieldValue	Dispositivo	Sokal-Sneath	0.7374
PathOrderFieldValue	Modelo	Euclídea	0.6196
PathOrderFieldValue	Modelo	Correlación	0.7954
PathOrderFieldValue	Modelo	Rogers-Tanimoto	0.7896
PathOrderFieldValue	Modelo	Sokal-Sneath	0.7374

Tabla IX
CONFIGURACIÓN DEL CONGLOMERADO JERÁRQUICO.

Parámetros	Valores
Universo	PathField
Métrica	Euclídea
Umbral	1.132
Criterio	Encadenamiento
Encadenamiento	Simple

IV-D. Resultados del Algoritmo de Conglomerado Jerárquico

Todas las ejecuciones se han completado con las diferentes configuraciones, mostradas en las Tablas VII y VIII, para cada uno de los conjuntos de datos utilizados en este trabajo. La Tabla IX muestra el resumen de las condiciones experimentales que se han utilizado en el algoritmo de conglomerados jerárquico.

IV-D1. Resultados obtenidos para la muestra de VISION:
Los resultados del experimento se muestran en la Figura 1 donde se puede ver que el número de conglomerados que han sido identificados, son 17 de los 13 modelos disponibles en el conjunto de datos de VISION. Los vídeos de las plataformas

YouTube y WhatsApp se han identificado casi por completo en un conglomerado cada uno. La marca Apple identifica varios conglomerados con un comportamiento diferente al de las otras marcas. Otros dispositivos de diferentes marcas,

IV-D2. Resultados Obtenidos para la Muestra de ACID: Como se puede ver en la Figura 3 hay 11 conglomerados de las 11 marcas que pertenecen al conjunto de datos. En este caso, a diferencia de la muestra de VISION, la marca Apple está correctamente clasificada en un solo grupo. Las marcas LG y Moto no pueden distinguirse con esta representación. En cuanto a las cámaras digitales, se observa que las marcas Canon y Olympus están correctamente clasificadas, pero no ocurre lo mismo con la marca Kodak, que no puede distinguirse de la marca Samsung.

La Figura 4 muestra gráficamente los conglomerados que se han generado según Modelo para la muestra de ACID y en la Tabla XI muestra los resultados numéricos asociados.

Todas las configuraciones mostradas en las Tablas VII y VIII se han utilizado para ejecutar OPTICS. En las ejecuciones se han fijado $minPoints=5$ y $\epsilon=0.01$ que son

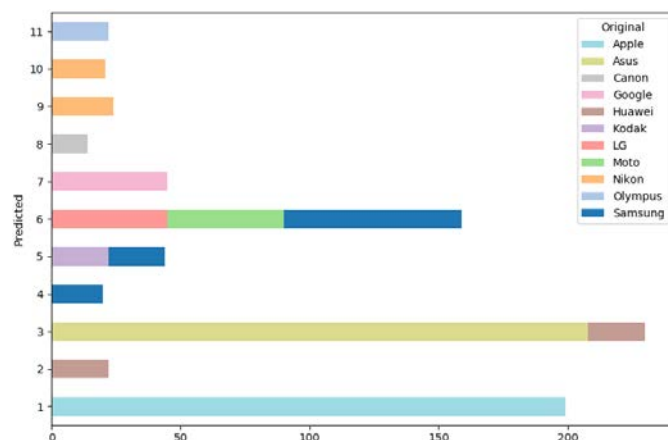


Tabla X
CONGLOMERADO JERÁROUICO AGRUPADOS POR MARCA.

Parámetro	VISION	ACID
#Marcas	13	11
#Conglomerados	17	11
RI	0.8839517587	0.8128426754
Homogeneidad	0.9195359995	0.8324380092
Compleitud	0.7970734665	0.8923993328

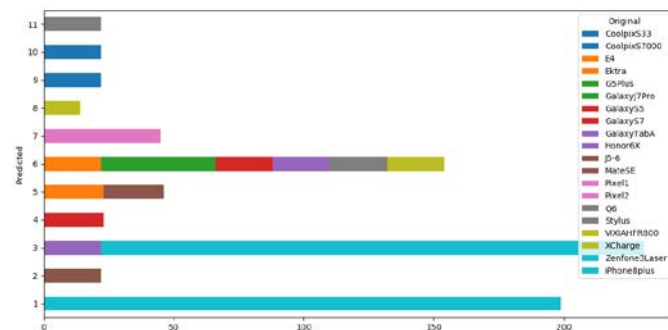


Tabla XI
PROCEDIMIENTO DE CONGLOMERADOS JERÁRQUICO POR MODELO
PARA LA MUESTRA DE ACID

#Modelos	20
#Conglomerados	11
RI	0.8233019504
Homogeneidad	0.778805082778
Completitud	1.0

IV-E1. Resultados para la Muestra de VISION: Como se puede ver en la Figura 5 el algoritmo ha generado 23 conglomerados de las 13 marcas que pertenecen al conjunto de datos de VISION.

92

La metodología propuesta ha sido validada a través de dos conjuntos de datos a los que ha sido aplicada con la misma selección de parámetros a fin de obtener resultados comparables. Los conjuntos de datos utilizados se han obtenido mediante el muestreo sobre las dos bases de datos más actuales de la literatura. Las bases de datos contienen vídeos de diversas tecnologías: vídeos nativos de dispositivos móviles, vídeos nativos de cámaras digitales y vídeos que se han descargado de las plataformas WhatsApp y YouTube de tal manera que se ha intentado obtener muestras suficientemente significativas para poder llevar a acabo el estudio.

La metodología propuesta es lo suficientemente general como para poder aplicarla y adaptarla a otro tipos de datos (modificación de la obtención preliminar de los mismos, modificación de la representación primitiva mediante la combinación de otros campos, etc.) , así como aplicar otras técnicas de clasificación presentes en Análisis Multivariante (técnicas de clasificación no jerárquica, utilización de métodos basados en modelos estadísticos [25], entre otros). Como se ha visto en la los resultados numéricos obtenidos de las muestras, los algoritmos de agrupación propuestos han proporcionado buenos resultados desde el punto de vista de la clasificación, obteniéndose precisiones superiores al 89 %.

AGRADECIMIENTOS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700326. This paper has also received funding from THEIA (Techniques for Integrity and authentication of multimedia files of mobile devices) UCM project (FEI-EU-19-04).



REFERENCIAS

- [1] D. R. Hayes, *A Practical Guide to Computer Forensics Investigations*. Pearson Education, August 2015.
- [2] M. C. Stamm, M. Wu, and K. Liu, "Information Forensics: An Overview of the First Decade," *IEEE Access*, vol. 1, pp. 167–200, March 2013.
- [3] P. Bestagini, M. Fontani, S. Milani, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, "An Overview on Video Forensics," in *Proceedings of the 20th European Signal Processing Conference*, Bucharest, Romania, August 2012, pp. 1229–1233.
- [4] B. Hosler, O. Mayer, B. Bayar, X. Zhao, C. Chen, J. A. Shackelford, and M. C. Stamm, "A Video Camera Model Identification System Using Deep Learning and Fusion," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, UK, May 2019, pp. 8271–8275.
- [5] B. Sevinc, H. Sencar, and N. Memon, "Classification of Digital Camera-Models Based on Demosaicing Artifacts," *Digital Investigation*, vol. 5, no. 1-2, pp. 49–59, September 2008.
- [6] H. Cao and A. C. Kot, "Accurate Detection of Demosaicing Regularity for Digital Image Forensics," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 899–910, December 2009.
- [7] X. Zhao and M. C. Stamm, "Computationally Efficient Demosaicing Filter Estimation for Forensic Camera Model Identification," in *Proceedings of the IEEE International Conference on Image Processing*, Phoenix, Arizona, USA, September 2016, pp. 151–155.
- [8] M. A. Qureshi, C.-H. Deriche, M. Choi, H.-Y. Lee, and H.-K. Lee, "Estimation of Color Modification in Digital Images by CFA Pattern Change," *Forensic Science International*, vol. 226, no. 1-3, pp. 94–105, March 2013.
- [9] F. M. Peng and D.-L. Zhou, "Discriminating Natural Images and Computer Generated Graphics Based on the Impact of CFA Interpolation on the Correlation of PRNU," *Digital Investigation*, vol. 11, no. 2, pp. 111–119, June 2014.
- [10] P. Mullan, C. Riess, and F. Freiling, "Forensic Source Identification Using JPEG Image Headers: The Case of Smartphones," *Digital Investigation*, vol. 28, no. Supplement, pp. S68–S76, April 2019.
- [11] J. Lukáš, J. Fridrich, and M. Goljan, "Determining Digital Image Origin Using Sensor Imperfections," in *Proceedings of the Image and Video Communications and Processing 2005*, San Jose, California, USA, March 2005, pp. 249–260.
- [12] —, "Detecting Digital Image Forgeries Using Sensor Pattern Noise," in *Proceedings of the SPIE-The International Society for Optical Engineering*, San Jose, California, USA, February 2006, pp. 362–372.
- [13] M. Chen, J. Fridrich, and M. Goljan, "Digital Imaging Sensor Identification (Further Study)," in *Proceedings of the SPIE Electronic Imaging, Security, Steganography, Watermarking of Multimedia Contents IX*, February 2007, pp. 1–14.
- [14] T. Filler, J. Fridrich, and M. Goljan, "Using Sensor Pattern Noise for Camera Model Identification," in *Proceedings of the 15th IEEE International Conference on Image Processing*, October 2008, pp. 1296–1299.
- [15] Y. Hu, B. Yu, and C. Jian, "Source Camera Identification Using Large Components of Sensor Pattern Noise," in *Proceedings of the Second International Conference on Computer Science and its Applications*, Jeju Island, Korea, January 2010, pp. 1–5.
- [16] C.-T. Li, "Source Camera Identification Using Enhanced Sensor Pattern Noise," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, June 2010.
- [17] L. J. García Villalba, A. L. Sandoval Orozco, R. Ramos López, and J. Hernández Castro, "Identification of Smartphone Brand and Model via Forensic Video Analysis," *Expert Systems with Applications*, vol. 55, pp. 59–69, August 2016.
- [18] "ISO/IEC 14496-12:2015: Coding of Audio-Visual Objects-Part 12: ISO base media file format," <https://www.iso.org/standard/68960.html>, 2017.
- [19] T. Gloe, A. Fischer, and M. Kirchner, "Forensic Analysis of Video File Formats," *Digital Investigation*, vol. 11, no. Supplement 1, pp. 68–76, May 2014.
- [20] M. Iuliani, D. Shullani, M. Fontani, S. Meucci, and A. Piva, "A Video Forensic Framework for the Unsupervised Analysis of MP4-Like File Container," *Institute of Electrical and Electronics Engineers Transactions on Information Forensics and Security*, vol. 14, no. 3, pp. 635–645, March 2019.
- [21] J. Song, K. Lee, Y. Lee, Wan, and H. Lee, "Integrity Verification of the Ordered data Structures in Manipulated Video Content," *Digital Investigation*, vol. 18, pp. 1–7, September 2016.
- [22] "QuickTime File Format Specification: Overview," <https://developer.apple.com/library/archive/documentation/QuickTime/QTFF/QTFFChap1/qtff1.html>, September 2016.
- [23] D. Shullani, M. Fontani, M. Iuliani, O. Alshaya, and A. Piva, "VISION: a Video and Image dataset for Source Identification," *EURASIP Journal on Information Security*, vol. 1, no. 15, pp. 1–16, October 2017.
- [24] M. Esteban Cobo, "Herramienta para la Extracción Automática de Metadatos en Vídeos de Dispositivos Móviles," Facultad de Informática, Universidad Complutense de Madrid, Spain, Trabajo de Fin de Grado, August 2016.
- [25] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Pearson Education Inc.(US), 2007.
- [26] M. Ester, H. Kriegel, J. Sander, and X. Xiaowei, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, September 1996, pp. 226–231.
- [27] M. Ankerst, M. Breunig, H. P. Kriegel, and J. Sander, "OPTICS: Ordering Points To Identify the Clustering Structure," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, New York, USA, June 1999, pp. 49–60.
- [28] S. Khan and T. Bianchi, "Fast Image Clustering Based on Camera Fingerprint Ordering," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, Shanghai, China, July 2019, pp. 766–771.
- [29] L. J. García Villalba, A. L. Sandoval Orozco, and J. Rosales Corripio, "Smartphone Image Clustering," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1927–1940, March 2015.
- [30] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, "Blind PRNU-Based Image Clustering for Source Identification," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2197–2211, September 2017.
- [31] N. X. Vinh, J. Epps, and J. Bailey, "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance," *Journal of Machine Learning Research*, vol. 11, pp. 2837–2854, October 2010.

Raquel Ramos López received his Computer Science Engineering degree and a M.S. degree in Computer Science, both from the Universidad Complutense of Madrid. She is currently a Ph.D. student of Computer Engineering in the GASS Group (Group of Analysis, Security and Systems) in Universidad Complutense de Madrid. Her main research interests are information security and its applications.

Elena Almaraz Luengo received a Mathematics degree from the University Complutense of Madrid in 2005, a Statistical Sciences and Techniques degree from the University Complutense of Madrid in 2007 and a Business and Administration degree from the National Distance Education University in 2015. She is Doctor in Mathematics from the University Complutense of Madrid since 2007 and holds a Master's Degree in Advanced Mathematics with specialization in Statistics and Operations Research, from the National Distance Education University in 2010. She is currently an Assistant Professor in the Department of Statistical and Operational Research in the Faculty of Mathematics Sciences of the University Complutense of Madrid. Her main interest are statistical techniques, probability, information security and applications.

Ana Lucila Sandoval Orozco received a Computer Science Engineering degree from the Universidad Autónoma del Caribe (Colombia) in 2001. She holds a Specialization Course in Computer Networks (2006) from the Universidad del Norte (Colombia), and holds a M.Sc. in Research in Computer Science (2009) and a Ph.D. in Computer Science (2014), both from the Universidad Complutense de Madrid (Spain). She is currently a postdoctoral researcher at Universidad Complutense de Madrid (Spain). Her main research interests are coding theory, information security and its applications.

Luis Javier García Villalba received a Telecommunication Engineering degree from the Universidad de Málaga (Spain) in 1993 and holds a Ph.D. in Computer Science (1999) from the Universidad Politécnica de Madrid (Spain). Visiting Scholar at COSIC (Computer Security and Industrial Cryptography, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, Belgium) in 2000 and Visiting Scientist at IBM Research Division (IBM Almaden Research Center, San Jose, CA, USA) in 2001 and 2002, he is currently Associate Professor of the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM) and Head of Complutense Research Group GASS (Group of Analysis, Security and Systems) which is located in the Faculty of Computer Science and Engineering at the UCM Campus. His professional experience includes the management of both national and international research projects and both public (Spanish Ministry of R&D, Spanish Ministry of Defence, Horizon 2020 - European Commission, ...) and private financing (Hitachi, IBM, Nokia, Safelayer Secure Communications, TB Solutions Security, ...). Author or co-author of numerous international publications is editor or guest editor of numerous journals such as Entropy, MDPI, Future Generation Computer Systems, Future Internet, MDPI, IEEE Latin America Transactions, IET Communications, IET Networks, IET Wireless Sensor Systems, International Journal of Ad Hoc and Ubiquitous Computing, International Journal of Multimedia and Ubiquitous Engineering (IJMUE), Journal of Supercomputing, Sensors, MDPI, etc.

ANÁLISIS DE AMENAZAS Y VULNERABILIDADES

Modelo SOCIA. Una reflexión conceptual y práctica desde la perspectiva del adversario

Jeimy J. Cano M.

Resumen— El objetivo de este artículo es presentar una propuesta metodológica denominada SOCIA para modelar al adversario y sus movimientos inesperados a partir de la incertidumbre, que incluyendo las reflexiones de los diferentes actores organizacionales, permite elaborar una visión diferente y complementaria para comprender mejor los retos y acciones de los atacantes. Para ello, se aplicó el instrumento desarrollado por cada uno de los elementos que constituyen la propuesta en el departamento de auditoría de una institución financiera, que organizado en grupos de trabajo interdisciplinarios, procedió a analizar varios casos de ciberataques en el sector bancario. Los resultados del ejercicio permiten construir una visión más elaborada y enriquecida de los riesgos actuales desde diferentes perspectivas para realizar simulaciones que desafían los mecanismos de seguridad y control actuales, y así poner de manifiesto posibles puntos ciegos en sus procesos y en la infraestructura tecnológica actual. En consecuencia, la SOCIA, crea un lenguaje común para elaborar una comprensión de la dinámica del adversario, no sólo a partir de su capacidad técnica, sino de sus técnicas y estrategias para configurar la acción no autorizada, aumentando la conciencia de la empresa sobre las capacidades del agresor y su impacto en la promesa de valor de la empresa.

Palabras clave — Adversario, Amenazas emergentes, Consciencia situacional, Estándares, Confianza

I. INTRODUCCIÓN

CON el pasar del tiempo, las prácticas de seguridad de la información han venido evolucionando tratando de establecer una nueva frontera de protección, que demore al atacante y trate de disuadirlo de sus acciones contrarias frente a los activos de información estratégicos de las organizaciones. En este ejercicio, se han actualizado las normas ISO, los marcos de trabajo disponibles en tecnología de información, así como los listados de controles que las empresas desarrollan para establecer su línea base de operaciones y sobre la cual, cada cierto tiempo se adelanta su validación de efectividad por parte de las áreas de auditoría de tecnología de información [1].

Si bien estas actividades, se han convertido en “*commodities*” de la práctica de seguridad de la información, el atacante no ha caído en esa realidad y por el contrario, ha avanzado en su capacidad de afectación y agresión, en niveles más sofisticados, menos visibles y más efectivos, adoptando las nuevas posibilidades tecnológicas y explorado nuevas vulnerabilidades del entorno [24], fruto de la mayor densidad digital disponible ahora en un escenario más hiperconectado y tecnológicamente modificado [2, 19].

En consecuencia, contar con prácticas estándares de seguridad de la información, si bien permite a las organizaciones asegurar un espacio de acción frente a las amenazas conocidas, deja poco margen para que puedan responder o anticipar, las propuestas inciertas de los agresores, que por lo general terminan sorprendiendo tanto a los especialistas en seguridad de la información, como a las diferentes áreas de las empresas [3, 25]. Lo anterior revela las limitaciones inherentes de los estándares conocidos, que implica una revisión en detalle de los nuevos entornos de operación de las empresas, mediados por redes sociales [30], para repensar nuevamente dichas prácticas y pensar ahora, en el desarrollo de capacidades distintivas para atender los retos de la protección de la información.

Dicho lo anterior, se hace necesario reflexionar sobre la manera de incluir la figura del adversario en los marcos de seguridad y control disponibles a la fecha, considerando que éste y sus técnicas son el nuevo reto a comprender, atender y superar, de tal forma, que no sólo sea la vista del analista tradicional de seguridad de la información, basado en un listado de controles el que atienda las realidades emergentes del contexto organizacional [26], sino que éste desarrolle una capacidad de análisis, prospectiva y aprendizaje que le permita caminar en terrenos algunas veces inciertos y en otros, inesperados, para tratar de descifrar las estrategias de los atacantes y así, proponer acciones alternativas de mitigación o contención, teniendo presente que en algún momento el agresor va a tener éxito [4].

De acuerdo con lo expresado previamente, este artículo presenta una propuesta conceptual y práctica para modelar el adversario. Su diseño busca complementar los análisis ya existentes de los entornos de las organizaciones, para que apalancados en una visión sistémica de riesgos (conocidos, latentes, focales y emergentes) [5], se puedan plantear escenarios de análisis prospectivos, basados en casos previamente documentados, experiencias propias de la industria o alternativas novedosas que los propios analistas de seguridad de la información u otros perfiles que puedan sugerir, con lo cual se habilita a la organización para anticipar amenazas emergentes que posiblemente no se hayan identificado previamente.

Por tanto, este documento para dar cuenta de sus reflexiones se organiza de la siguiente manera: en la sección II se presenta una revisión y análisis de algunas propuestas actuales sobre el modelaje de amenazas; en la sección III se introduce el

modelo SOCIA como una propuesta conceptual que sugiere una vista ágil e inclusiva del adversario. En la sección IV se presenta y explica el uso del modelo en un caso particular utilizado en un taller en una entidad financiera, y finalmente en la sección V se plantean algunas conclusiones sobre el modelo presentado y su aplicación.

II. MODELANDO LAS AMENAZAS. REVISIÓN DE ALGUNAS PROPUESTAS VIGENTES

Entender al adversario no es un tema nuevo en seguridad de la información. Es una dinámica que en muchas ocasiones se usa para tratar de advertir alguna violación de un control o acción adversa de un individuo frente a la protección de un activo. En este sentido, los profesionales de seguridad de la información encuentran en la forma de pensar del atacante, una vista de aquello que puede salir distinto a lo planeado, y poder ver cómo es posible superar o afectar la medida de protección previamente instalada y probada en el sistema o activo protegido [23].

Sin perjuicio de lo anterior, y considerando la inestabilidad del entorno de operaciones actual de las organizaciones, se han venido desarrollando propuestas alternativas que buscan modelar las actividades y tácticas de los atacantes, tratando de establecer algunos parámetros de análisis detallados, generalmente basados en ataques previos documentados, con el fin de establecer un perfil del agresor que pueda ser visualizado desde las estrategias utilizadas y las acciones técnicas realizadas, para configurar una “firma” del adversario que ilustre a los analistas cómo es su modo de operación y así advertir, las acciones de mitigación o contención necesarias de forma anticipada [6].

Dentro del listado de propuestas de modelaje de amenazas se encuentran las siguientes: [22]

- STRIDE – Inventado por Loren Kohnfelder y Praerit Garg en 1999 y adoptado por Microsoft en 2002. Ha evolucionado con el tiempo incluyendo las variantes STRIDE por elemento y STRIDE por interacción.

- P.A.S.T.A (*Process for Attack Simulation and Threat Analysis*) - es un marco de modelado de amenazas centrado en el riesgo desarrollado en 2012 por Tony Uceda Vélez. Contiene siete etapas, cada una con múltiples actividades.

- *Attack Trees* - Desarrollado por Bruce Schneider en 1999, se aplicó inicialmente como método propio. A la fecha se ha combinado con otros métodos y marcos.

- *CyberKill Chain* – Desarrollado por la firma Lockheed Martin en 2011. Es una metodología aceptada por la industria para entender cómo un atacante llevará a cabo las actividades necesarias para causar daño a una organización.

- MITRE ATT&CK (Tácticas, Técnicas y Conocimiento Común de Adversarios) – Desarrollado por el MITRE² en

2013. Es una base de conocimientos de acceso global sobre tácticas y técnicas adversarias basadas en observaciones del mundo real.

A continuación, se detallan cada una de ellas.

A. STRIDE

STRIDE, es el acrónimo de seis amenazas claves (en inglés) que establece este método para modelar las posibles acciones de los adversarios: *spoofing*, *tampering*, *repudiation*, *information disclosure*, *denial of service* y *elevation of privilege*, que traducido al español se leería como suplantar, alterar, repudiar, revelar información, denegar el servicio, y elevar privilegios. Esta propuesta puede ser utilizada desde múltiples enfoques: como una forma para pensar como el atacante y establecer maneras de vulnerar el sistema; como una forma de analizar los elementos del sistema y sus vulnerabilidades o como una estrategia para atacar las interacciones que tiene el objeto bajo evaluación con sus otros componentes [7].

Con este método, los analistas o personas participantes de la revisión del sistema bajo evaluación procuran establecer aquellas cosas que pueden salir mal con el sistema, los momentos o acciones que se pueden materializar y afectar el adecuado funcionamiento del mismo. De igual forma, por cada amenaza enumerada se tiene el principio y/o servicio de seguridad que se requiere asegurar indicado entre paréntesis: *spoofing* (autenticación), *tampering* (integridad), *repudiation* (no repudio), *information disclosure* (confidencialidad), *denial of service* (disponibilidad) y *elevation of privilege* (autorización) [6].

Adelantar un ejercicio siguiendo lo sugerido por STRIDE, permite tener un escenario base de amenazas en los diferentes puntos del sistema objetivo, los cuales son insumo para entender los impactos de su malfuncionamiento, uso inadecuado o falla, así como para plantear los mecanismos de seguridad y control requeridos para hacerlo más resistente a los ataques. El método prevé algunos controles para cada uno de los principios y servicios de seguridad asociados con las amenazas.

B. P.A.S.T.A

El proceso de simulación de ataques y análisis de amenazas establece siete (7) pasos metodológicos claves para establecer el marco de acción frente a situaciones inciertas del sistema o componente que se analiza: [8]

- *Definir los objetivos* – Este primer paso incluye detallar los objetivos de negocio que desarrolla el sistema, enumerar los requerimientos de seguridad y cumplimiento necesarios y el análisis de impacto en el negocio si este componente o pieza de software falla.

- *Definir el alcance técnico* – Esta fase demanda establecer los límites técnicos de la solución bajo análisis, con el fin de identificar los recursos tangibles que pueden destinarse a la realización de una amenaza determinada. Con esta

² La Corporación MITRE fue constituida en 1958 como una compañía privada sin fines de lucro para proporcionar orientación técnica y de ingeniería al gobierno federal norteamericano. Fuente: <https://www.mitre.org/about/our-history>

información se puede tener una mayor claridad en la comprensión de las oportunidades que un atacante puede tener contra los activos de un entorno objetivo.

- *Descomponer el sistema* – En este paso el sistema analizado se divide en componentes individuales que pueden ser atacados de forma única por un atacante. Este ejercicio permite a un adversario descomponer el sistema objetivo en partes para identificar la mejor forma para atacarlo.

- *Análisis de amenazas* – Esta parte del modelo, considerando la etapa anterior, establecen las razones del por qué y cómo se pueden realizar los ataques al sistema. Para ello se desarrollan preguntas orientadoras que ayuden a visualizar los motivos que hay detrás del ataque al entorno del sistema.

- *Conocimiento de vulnerabilidades* – En este punto del proceso, considerando la información recabada en los pasos previos, se identifican las fallas de diseño o funcionalidad de cada uno de los componentes del sistema, ejercicio que requiere experiencia técnica y experticia práctica que permita identificar las vulnerabilidades propias de sistema objetivo y sus relaciones.

- *Construir el árbol de ataques* – El objetivo de esta etapa se centra en determinar la viabilidad de materializar las vulnerabilidades identificadas en el paso anterior. El éxito de este paso está en la capacidad del profesional que adelanta las pruebas de vulneración para demostrar que las debilidades identificadas son realmente explotables.

- *Análisis de riesgos e impactos* – Esta última fase el modelador de amenazas toma los resultados de la fase anterior y cuantifica los índices de probabilidad de éxito en algunos de los ataques definidos, con el fin de establecer el nivel de riesgo de exposición que tiene el sistema, así como los impactos en sus objetivos de negocio definidos previamente.

Adelantar un ejercicio de análisis de amenazas y vulnerabilidades de un sistema de información o aplicación utilizando P.A.S.T.A, ofrece una visual detallada de sus interacciones, las implicaciones técnicas y de negocio que son claves tener en cuenta y los impactos derivados de la posible materialización de un evento no deseado. En esta dinámica deben participar equipos de trabajo de las áreas afectadas y los profesionales de seguridad con el fin de establecer una vista común de los retos que implica el sistema y sus posibles fallas de identificadas.

C. Attack Trees

Los árboles de ataque son esencialmente diagramas que representan ataques a un sistema en forma de árbol. La raíz del árbol es el objetivo del ataque, y las hojas son las formas de lograr ese objetivo. Cada meta se representa como un árbol separado. Por lo tanto, el análisis de amenazas del sistema produce un conjunto de árboles de ataque. Para tener diferentes opciones de ataque se incorporan nodos “AND” para representar los pasos requeridos para concretar el ataque, y los “OR” para indicar alternativas para conseguir el resultado esperado por el agresor [9].

Adelantar un análisis con esta propuesta metodológica, demanda un conocimiento detallado del sistema que se analiza, dado que no se ofrece mayor detalle sobre submetas, ataques o riesgos que se pueden materializar. Por lo general, esta estrategia de modelaje de amenazas se usa en conjunto con otras, con el fin de poder comprender con detalle la interacción del sistema que se analiza y las implicaciones de seguridad de sus diferentes componentes.

D. Cyberkill Chain

Esta propuesta de modelaje del adversario consta de siete pasos en los cuales se analiza tanto la postura del adversario como la del defensor. De lo anterior, resulta un ejercicio complementario para establecer posibles marcos de actuación que se pueden resolver bien, desde la implementación o actualización de tecnología, pasando por la gestión de los procesos de negocio y su seguridad, la revisión de la cultura organizacional o incluso, desde la generación de escenarios posibles que reten las condiciones vigentes de seguridad y control de la organización.

Las etapas son: [10]

1. Reconocimiento: exploración intencional del objetivo para identificar posibles vulnerabilidades.

2. Militarización: el adversario crea código malicioso o condiciones particulares adaptadas a las debilidades de los objetivos.

3. Entrega: el adversario ubica el código malicioso o configura el escenario en el objetivo.

4. Explotación: se ejecuta el código malicioso o se materializa el escenario estudiado según el plan del adversario.

5. Instalación: la condición adversa se enraíza y conecta en el contexto del objetivo.

6. Comando y control: el código malicioso se comunica con los sistemas del adversario y se crea una condición de control remoto.

7. Acción sobre los objetivos: el código malicioso realiza la acción planeada o el escenario desarrolla y materializa los efectos deseados.

Asumir un enfoque desde el adversario, como la cadena *Cyber Kill*, permite no solo retar los conocimientos y prácticas estándares generalmente adoptadas en las organizaciones, sino que aumenta su capacidad para entender y responder de manera dinámica a las acciones no autorizadas de terceros que quieran resquebrajar la confianza de sus clientes en su entorno de operaciones y desarrollo de actividad comercial.

E. MITRE ATT&CK

De manera general se puede expresar que la propuesta ATT&CK es un modelo de comportamiento que consiste en los siguientes componentes básicos: [11]

- Táctica, que denota objetivos tácticos de los adversarios a corto plazo durante un ataque.

- Técnicas, donde se describen los medios por los cuales los adversarios logran sus metas tácticas.

- Los usos documentados de las técnicas de los

adversarios, y otros metadatos relevantes asociados con las técnicas.

Este modelo se puede utilizar para: [11]

- *Simular al adversario* – Establecer escenarios para probar y verificar las contramedidas instaladas en las organizaciones, frente a técnicas concretas de los adversarios y sus posibles impactos.

- *Aplicar la lógica del adversario* – Asumir el papel de un agresor sin utilizar información de amenazas conocidas, con el fin de desarrollar una operación sin ser detectado, para mostrar el impacto de una operación exitosa y evaluar la capacidad de detección y reacción de una organización frente a estos eventos.

- *Desarrollar analítica de comportamiento* - Usarlo como una herramienta para construir y probar análisis de comportamiento para detectar actividad potencialmente maliciosa dentro de un ambiente y establecer acciones concretas frente a las mismas.

- *Valorar brechas defensivas* – Considerando un modelo de adversario concreto centrado en el comportamiento, validar las herramientas, el monitoreo y mitigación de brechas en las medidas defensivas existentes dentro de la empresa de una organización. Las brechas identificadas configuran una forma de priorizar las inversiones para la mejora del programa de seguridad.

- *Enriquecer la inteligencia de amenazas* – Permite a la organización entender y documentar los perfiles de los grupos adversarios desde una perspectiva de comportamiento que es agnóstica de las herramientas que el grupo pueda utilizar.

El uso de cualquiera de estas propuestas para modelar amenazas y adversarios establece un marco general de acción que permite a los analistas de seguridad de la información ir más allá de la vista tradicional de controles y abrirse a un ejercicio más detallado y relacional, que lo habilita para comprender mejor los riesgos emergentes en sus entornos particulares. Las diferencias entre ellas se encuentran en la aplicación de estas, las cuales demandan esfuerzos particularmente intensivos en dedicación y recursos, con el fin de lograr el objetivo deseado.

Lo anterior implica que seleccionar alguno de ellos, exige conocer la agilidad de la solución requerida, sus tiempos de despliegue, los diferentes equipos de trabajo involucrados, así como el apetito de riesgo corporativo establecido por la alta gerencia, con el fin de identificar el enfoque que mejor se ajusta a la necesidad de la organización. A manera de resumen podemos detallar en la tabla 1, cada una de estas propuestas con sus características y limitaciones.

Tabla 1.

Resumen de características y limitaciones de las propuesta para modelar amenazas

Propuesta metodológica	Características	Limitaciones
STRIDE	• Fácil de usar • Focaliza la estrategia de mitigación • Uno de los más usados	* Demanda alto consumo de tiempo y recursos * Requiere conocimiento detallado del sistema a modelar
P.A.S.T.A	• Focaliza la estrategia de mitigación • Motiva la colaboración entre los participantes • Facilita el plan priorizado de mitigación	* Demandante en los detalles del sistema que se analiza * Exige documentación detallada de las partes del sistema y sus riesgos
Attack trees	• Permite ver con claridad las posibles fallas del sistema • Fácil de usar • Identifica las estrategias de mitigación relevantes	* Requiere conocimiento detallado del sistema para su análisis * No ofrece orientación sobre riesgos, fallas o problemas a analizar
Cyberkill Chain	• Permite comprender los pasos del atacante • Establece una vista del adversario y sus acciones • Habilita a la organización para reconocer actividades maliciosas	* Requiere de conocimiento experto * Se pueden presentar conflictos de lenguaje entre especialistas técnicos de los sistemas bajo análisis y las personas de las áreas de negocio
MITRE ATT&CK	• Detalla el comportamiento del atacante • Establece con claridad las técnicas y tácticas que puede usar • Basado en conocimiento cierto de eventos y amenazas materializadas	* Toma tiempo la caracterización del perfil requerido * Asume un conocimiento técnico requerido para comprender el patrón del adversario

Fuente: Elaboración propia.

III. MODELO SOCIA. UNA PROPUESTA CONCEPTUAL PARA MODELAR EL ADVERSARIO

Los modelos previamente revisados establecen una serie de condiciones y características que por lo general demandan pasos detallados y procedimientos específicos para lograr los

objetivos trazados en el modelamiento de las amenazas y de los adversarios. En este sentido, aplicar cualquiera de ellos, exige una planeación detallada, entregables concretos y tiempos definidos para lograr esa visión novedosa y emergente que revele los puntos ciegos en la evaluación de vulnerabilidades y brechas de seguridad.

En este entendido, se hace necesario contar con una herramienta liviana, sencilla, efectiva y funcional, que sin pretender desarrollar los alcances que tienen las propuestas analizadas previamente, si ofrezca una forma ágil de lograr un ejercicio base que permita entender el adversario, no como un elemento abierto en un escenario, sino como una realidad evidente y tangible dentro de la dinámica de los procesos, con un mínimo de variables a revisar. Considerando lo anterior, es posible plantear una reflexión enriquecida con los participantes claves del proceso de negocio y los sistemas de información que se quieren analizar.

Siguiendo las reflexiones de Do, Martini y Raymond Choo [12] relacionadas con el papel del adversario en las investigaciones aplicadas en seguridad, y las consideraciones del *Common Sense Security Framework*³ (CSSF), se han identificado cinco (5) variables claves para desarrollar una propuesta metodológica para modelar el atacante denominada SOCIA. SOCIA es una apuesta conceptual de uso ágil para utilizar en el contexto corporativo, con el fin de crear conciencia situacional [13] y reflexiones prospectivas sobre posibles adversarios que pueden tener las organizaciones. Para lograrlo, se parte de la revisión de situaciones previamente materializadas tanto a nivel nacional como internacional, que permitan una visión sistémica de los retos que plantea un atacante, y la participación de diferentes perfiles de colaboradores en la empresa [14,17].

SOCIA es el acrónimo que vincula cinco variables, a saber: Supuestos, Objetivos, Capacidades, Impactos y Aprendizajes (figura 1). A continuación, se detallan cada una de estas variables que configuran una forma de modelar posibles adversarios en el contexto de las organizaciones modernas, considerando la estrategia metodológica del CSSF, de utilizar tres (3) elementos claves para contextualizar el tema, así como sus respectivas preguntas orientadoras.



Figura 1. Modelo SOCIA. Fuente: Elaboración propia

A. Supuestos

Esta primera variable establece las consideraciones básicas de las condiciones y perfil de adversario. En este punto se busca conocer un poco el contexto donde el adversario se mueve, sus conocimientos y recursos disponibles, y para ello, se establecen los siguientes tres (3) elementos con sus preguntas claves:

- Entorno: Interno/Externo - ¿Dónde puede estar ubicado el adversario: dentro o fuera?
- Recursos: Accesos/Equipo - ¿Tiene accesos disponibles bien sea físicos o lógicos, y cuenta con el equipo necesario para lograr sus propósitos?
- Conocimiento: Alto/Medio/Bajo - ¿El adversario cuenta con los conocimientos técnicos y prácticos requeridos para concretar sus acciones no autorizadas?

B. Objetivos

Esta segunda variable busca conocer las motivaciones y datos que son de interés del adversario. En este momento se indaga sobre qué tipo de datos son realmente de interés por parte del atacante, y ver si es posible su monetización. Es claro que no siempre el agresor busca un lucro económico, sino que sus motivaciones pueden ser políticas o ideológicas. Los elementos claves y sus preguntas son:

- Motivación: ¿Por qué? - ¿Cuál es la motivación del adversario para atacar a la organización?
- Activo: ¿Qué tipo de datos? - ¿Qué tipo de datos pueden ser de su interés?
- Monetización: ¿Cuánto puede costar? - ¿Esta información se puede monetizar rápidamente en el mercado?

C. Capacidades

Esta tercera variable revela los patrones de aprendizaje/desaprendizaje que el atacante tiene para producir un daño. En este segmento se evalúa la acción del agresor, cómo se apalanca para concretar sus acciones, y que tipo de afectación ha logrado. Para ello, se detallan a continuación los elementos claves y sus preguntas:

³ Detalles en: <https://commonsenseframework.org/>

- Acción: Pasiva/Activa - ¿La actuación del adversario será pasiva o activa?
- Plataforma: Local/Ecosistema - ¿Sus acciones estarán apalancadas localmente o con apoyo de un ecosistema público?
- Afectación: Baja/Media/Alta - ¿El nivel de afectación que puede generar es alto, medio o bajo?

D. Impactos

Esta cuarta variable revisa las consecuencias concretas que el adversario puede ocasionar con sus capacidades. Es un ejercicio de mirar hacia el interior de la empresa donde se pueden ocasionar las perturbaciones y pérdidas que afecten la organización a nivel del negocio, los procesos y la infraestructura. Los elementos claves y sus preguntas orientadoras son:

- Negocio: Parcial/Total - ¿El negocio puede sufrir afectaciones parciales o totales?
- Procesos: Parcial/Total - ¿Los procesos pueden ser afectados de forma parcial o total?
- Infraestructura: Parcial/Total - ¿La infraestructura se puede comprometer parcial o totalmente?

E. Aprendizajes

La quinta y última variable explora las lecciones aprendidas, las acciones de remediación y la actualización de la postura de seguridad que la organización debe incorporar, para ajustar sus mecanismos de defensa y tratar de anticipar nuevos eventos en el futuro. Para ello, se detallan a renglón seguido los elementos claves y sus preguntas:

- Lecciones: Aprendidas/por aprender - Frente a un atacante con estas características, ¿qué lecciones por aprender tenemos?
- Remediación: Negocio/Proceso/Infraestructura - ¿Tenemos claros cómo deben ser los procesos de contención y remediación del negocio, el proceso y la infraestructura?
- Postura: Defender/Anticipar - ¿Cómo debe cambiar la postura de seguridad de la organización: defender o anticipar?

Estas cinco (5) variables, si bien no buscan ser exhaustivas ni detalladas en su desarrollo, sí facilitan una conversación abierta e incluyente desde los diferentes actores de la empresa, con el fin de revelar posibles puntos ciegos en su dinámica de la seguridad y ciberseguridad, a fin de que se posibilite la generación de una consciencia colectiva sobre los posibles adversarios y la manera como pueden afectarla; y de la misma forma, ver estrategias de defensa y anticipación distintas, que enriquezcan la batería disponible de estrategias de protección de los activos de información críticos de las organizaciones [27].

IV. APLICACIÓN DEL MODELO SOCIA. UNA VISIÓN PRÁCTICA DE SU USO

El Modelo SOCIA recientemente se utilizó en un ejercicio académico de revisión y análisis del escenario de amenazas

digitales emergentes de una entidad financiera en Colombia, realizado con los profesionales del área de la auditoría interna. En el desarrollo del taller de aplicación los auditores con diferentes profesiones como Contadores Públicos, Ingenieros de Sistemas, Administradores de Empresas, entre otros, se distribuyeron en tres grupos de trabajo interdisciplinarios con el fin de estudiar tres (3) casos de ciberataques internacionales. El objetivo fue conceptualizar las posturas de los atacantes y concretar reflexiones prácticas al interior de la dinámica de la entidad y su factibilidad, las cuales socializaron al final del ejercicio.

Para ello, se diseñó un formato de trabajo (ver figura 2), que integra las variables del modelo y sus elementos asociados, en el que se recogieron, con el apoyo del facilitador del ejercicio, las inquietudes y conversaciones que surgieron en cada equipo de trabajo durante el estudio del caso, cuya información fue proporcionada y complementada mediante búsquedas en internet.

SOCIA	Nombre del Caso:
Supuestos (Configura el adversario y su entorno)	ENTORNO: RECURSOS: CONOCIMIENTO:
Objetivos (Establece las motivaciones del adversario)	MOTIVACIÓN: ACTIVO: MONETIZACIÓN:
Capacidades (Identifica el nivel de daño que puede causar)	ACCIÓN: PLATAFORMA: AFECTACIÓN:
Impactos (Analiza las afectaciones claves que puede ocasionar)	NEGOCIO: PROCESOS: INFRAESTRUCTURA:
Aprendizajes (Revisa algunos puntos ciegos en el modelo de seguridad de la empresa)	LECCIONES: REMEDIACIÓN: POSTURA:

Figura 2. SOCIA. Formato de trabajo. Fuente: Elaboración propia.

Uno de los casos revisados durante el ejercicio, fue el ciberataque realizado recientemente al Banco Central de México – BANXICO, en el cual terceros no identificados y altamente documentados en el sistema de pagos interbancarios – SPEI, lograron desarrollar una acción coordinada con diferentes actores para robar dinero de bancos comerciales en este país.

A manera de ejemplo se presenta a continuación el resultado obtenido por uno de los equipos de trabajo (ver figura 3), así como algunas de sus reflexiones realizadas una vez concluida su presentación sobre el caso, siguiendo los elementos propios del Modelo SOCIA.

SOCIA	BANXICO
Supuestos (Configura el adversario y su entorno)	ENTORNO: Adversario Externo RECURSOS: Cuenta con equipo necesario para actuar y apoyo para desplegar la acción no autorizada. CONOCIMIENTO: El adversario tiene un nivel de conocimiento especializado del sistema de pagos
Objetivos (Establece las motivaciones del adversario)	MOTIVACIÓN: Económica ACTIVO: Cuentas de bancos y datos de correntistas MONETIZACIÓN: Tanto como las transferencias no autorizadas que se logren.
Capacidades (Identifica el nivel de daño que puede causar)	ACCIÓN: Activa sobre el sistema de pagos PLATAFORMA: Desarrollo local AFECTACIÓN: Alta
Impactos (Analiza las afectaciones claves que puede ocasionar)	NEGOCIO: Parcial PROCESOS: Total INFRAESTRUCTURA: Parcial
Aprendizajes (Revela algunos puntos ciegos en el modelo de seguridad de la empresa)	LECCIONES: Acciones coordinadas con los bancos/Mejora de la gestión de incidentes REMEDIACIÓN: Análisis de vulnerabilidades / Pruebas conjuntas POSTURA: Supera la postura preventiva.

Figura 3. SOCIA. Resumen del resultado - Caso BANXICO. Fuente: Elaboración propia

Luego de presentar este análisis, los diferentes miembros del equipo hicieron comentarios y afirmaciones sobre lo revisado que se pueden resumir como sigue:

- Es posible que podamos tener un atacante como estos y, por tanto, nos exige un ejercicio de coordinación con los supervisores del sector, así como con el Banco Central.
- Es necesario adelantar simulaciones y ejercicios de este tipo, para prepararnos, junto con el sector en general, lo cual demanda enriquecer la postura de seguridad y control actual de la entidad.
- No podemos confiarnos en los mecanismos de seguridad y control que tenemos instalados en este momento, es necesario crear condiciones adversas dentro de la plataforma actual, para ver cómo reacciona y qué puntos ciegos tenemos a la fecha que son relevantes para asegurar la operación de la institución.

De otra parte, los equipos interdisciplinarios que se conformaron para desarrollar el ejercicio utilizando SOCIA reconocieron puntos de vista distintos, creando nuevas oportunidades de actividades conjuntas, con estrategias complementarias, que superan las aproximaciones exclusivamente tecnológicas que se tienen a la fecha para adelantar los planes de auditoría cuando se trata de los temas de seguridad y control en la entidad.

Estas reflexiones y otras adicionales, como resultado de los otros equipos conformados, establecieron referentes distintos de la mirada actual de la auditoría, considerando elementos que no estaban vinculados a la fecha en su plan de trabajo. Adicionalmente, fue un ejercicio para retar sus prácticas actuales basadas en estándares y marcos de trabajo reconocidos, donde se hicieron evidentes las limitaciones inherentes de los mismos respecto de entornos inciertos e inestables como los que actualmente tienen las organizaciones modernas.

V. CONCLUSIONES

En el ejercicio cotidiano de proteger y asegurar las organizaciones, las prácticas de seguridad y control guiadas por los estándares conocidos ha sido la norma sobre la cual, tanto las áreas de seguridad/ciberseguridad, como las de auditoría, han fundado sus diferentes acciones para generar un

marco de confianza regulado cimentado sobre controles. Bajo este entendido, las empresas han logrado movilizar iniciativas de negocio innovadoras que habilitan a sus clientes, para lanzarse a concretar una interacción cada vez más digital y con menos intervención humana [15, 20].

No obstante lo anterior, la evolución de los adversarios y sus inesperadas formas de concretar sus acciones adversas genera tensiones sobre los estándares de seguridad y control vigentes, creando una zona de incertidumbre donde es necesario repensar la manera como se entiende ahora la protección de la información en las empresas [16, 18]. Esto es, habilitar espacios de reflexión basados en los inciertos e inestabilidades del entorno, con la participación de diferentes perfiles de personas en las organizaciones, para encontrar puntos de vista diferentes a las certezas y definiciones que ofrecen las buenas prácticas internacionales.

Incluir al adversario como un nuevo elemento dentro de las reflexiones de protección de los activos digitales de las empresas, abre una ventana de aprendizaje y desaprendizaje, que busca crear espacios complementarios de discusión, que permitan contrastar las lecturas de los riesgos conocidos provistas por los estándares, frente a las posibilidades y escenarios alternativos que ofrece una propuesta metodológica como SOCIA, donde el agresor, lejos de ser un desconocido para el análisis, se convierte en materia de estudio para crear una consciencia situacional sobre posibles puntos ciegos que actualmente pueden estar vigentes en la dinámica de la empresa [28].

Vincular al atacante como materia de estudio, no es una invitación ni sugerencia para que las áreas de seguridad/ciberseguridad, o auditoría, se conviertan en los mismos verdugos de las organizaciones, sino una manera distinta para identificar los pasos de los agresores (bien sean internos o externos), caminar en la dinámica de su mente y motivaciones, con el fin de defender y anticipar acciones concretas que apoyen su preparación frente a eventos inciertos o inesperados, que permanecen latentes en la ejecución de sus procesos actuales sin signos de manifestación concretos [29].

SOCIA como propuesta metodológica se basa en la dinámica social y de negocio de la organización, que sin entrar en los detalles tecnológicos del ataque analizado, permite conectar diversas vistas de los participantes, creando un lenguaje común para elaborar un entendimiento de la dinámica del adversario, no desde su habilidad técnica exclusivamente, sino desde sus técnicas y estrategias para configurar la acción no autorizada, aumentando la sensibilización de las personas sobre las capacidades de los agresores y sus impactos en la promesa de valor de la compañía.

En consecuencia, SOCIA establece una forma distinta e incluyente para visualizar y caracterizar los posibles adversarios que puede tener una empresa y en este ejercicio,

comenzar a construir nuevas capacidades de defensa y anticipación, que permita a las organizaciones desarrollar escenarios, prototipos y simulaciones [21] que la ubiquen en una posición menos reactiva frente a las situaciones imprevistas, y la proyecten como una comunidad resiliente, experimentada en la atención de incidentes, que sabe que el adversario es inteligente y hábil, y que tarde o temprano tendrá éxito.

AGRADECIMIENTOS

El autor agradece a los doctores Jesús Vásquez Gómez, Especialista de Tecnología de Información del Banco Central de México y Gabriela María Saucedo Meza, Docente de la Universidad Militar Nueva Granada de Colombia, por sus valiosos y acertados comentarios que permitieron afinar las reflexiones de este artículo.

REFERENCIAS

- [1] Donaldson, S., Siegel, S., Williams, C. & Aslam, A. (2015). *Enterprise Security. How to build a successful cyberdefense program against advanced threats*. New York, USA: Apress
- [2] Sieber, S. & Zamora, J. (2018). The Cybersecurity Challenge in a High Digital Density World. *European Business Review*. Recuperado de: <https://www.europeanbusinessreview.com/the-cybersecurity-challenge-in-a-high-digital-density-world/>
- [3] Cano, J. (2019). Analyst and Adversary. Deconstructing the “Imaginary” of Security and Cybersecurity Professionals. *ISACA Journal*. 4. 48-53.
- [4] Tounsi, W. & Rais, H. (2017). A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & Security*, 72. 212-233. doi:10.1016/j.cose.2017.09.001.
- [5] Cano, J. (2017). The AREM Window: A Strategy to Anticipate Risk and Threats to Enterprise Cyber Security. *ISACA Journal*. 5.
- [6] Shostack, A. (2014). *Threat Modeling: Designing for Security*. Indianapolis, IN. USA: John Wiley & Sons.
- [7] Khan, R., McLaughlin, K., Laverty, D., & Sezer, S. (2018). STRIDE-based Threat Modeling for Cyber-Physical Systems. In 2017 IEEE PES: Innovative Smart Grid Technologies Conference Europe (ISGT-Europe): *Proceedings IEEE*. doi: 10.1109/ISGTEurope.2017.8260283
- [8] UcedaVélez, T. & Morana, M. (2015). *Risk Centric Threat Modeling: Process for Attack Simulation and Threat Analysis*. Hoboken, NJ. USA: John Wiley & Sons.
- [9] Schneier, B. (1999). Attack trees. Modeling security threats. *Dr. Dobbs Journal*. December. Recuperado de: https://www.schneier.com/academic/archives/1999/12/attack_trees.html
- [10] Hutchins, E., Cloppert, M. & Amin, R. (2011). Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains. *White Paper*. Lockheed Martin. Recuperado de: <https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf>
- [11] Strom, B., Applebaum, A., Miller, D., Nickels, K., Pennington, A. & Thomas, C. (2018). MITRE ATT&CK™: Design and Philosophy. *MITRE Product*. Recuperado de: <https://www.mitre.org/publications/technical-papers/mitre-attack-design-and-philosophy>
- [12] Do, Q., Martini, B., & Raymond Choo, K. K. (2018). The Role of the Adversary Model in Applied Security Research. *Computers & Security*. 81, 156-181. doi:10.1016/j.cose.2018.12.002
- [13] Franke, U. & Brynielsson, J. (2014). Cyber situational awareness – a systematic review of the literature, *Computers & Security*. 46. 18-31. doi: 10.1016/j.cose.2014.06.008.
- [14] Cano, J. (2019). Ciberriesgo. Aprendizaje de un riesgo sistémico, emergente y disruptivo. *Revista SISTEMAS*. Asociación Colombiana de Ingenieros de Sistemas. 63-73. Doi: 10.29236/sistemas.n151a5
- [15] Accenture (2019). Cyber resilience for industry X.0 in Europe. Agile cybersecurity in the age of the Industrial Internet of Things. Recuperado de: <https://accntu.re/2TvkQYy>
- [16] Campbell, N. (2017). Cyber Security Is A Business Risk, Not Just An IT Problem. *Forbes*. Recuperado de: <https://www.forbes.com/sites/edelmantechnology/2017/10/11/cyber-security-is-a-business-risk-not-just-an-it-problem/>
- [17] Carlton, M., Levy, J. & Ramim, M. (2019). Mitigating cyber attacks through the measurement of non-IT professionals’ cybersecurity skills. *Information & Computer Security*, 27(1), 101-121. Doi: 10.1108/ICS-11-2016-0088
- [18] Raban, Y. & Hauptman, A. (2018). Foresight of cyber security threat drivers and affecting technologies. *Foresight*. Doi: 10.1108/FS-02-2018-0020
- [19] Mossburg, E., Gelinne, J. & Calzada, H. (2017). Beneath the Surface of a cyberattack. A deeper look at business impact. *Deloitte*. Recuperado de: <https://www2.deloitte.com/us/en/pages/risk/articles/hidden-business-impact-of-cyberattack.html>
- [20] Weill, P., Apel, T., Woerner, S. & Banner, J. (2019). It Pays to Have a Digitally Savvy Board. *Sloan Management Review*. Recuperado de: <https://sloanreview.mit.edu/article/it-pays-to-have-a-digitally-savvy-board/>
- [21] Cano, J. (2018). Repensando los fundamentos de la gestión de riesgos. Una propuesta conceptual desde la incertidumbre y la complejidad. *Revista Ibérica de Tecnología y Sistemas de la Información*. E15. Abril. 76-87.
- [22] Shevchenko, N., Chick, T., O’Riordan, P., Scanlon, T. & Woody, C. (2018). Threat modeling: a summary of available methods. *SEI Carnegie Mellon University*. Recuperado de: https://resources.sei.cmu.edu/asset_files/WhitePaper/2018_019_001_524597.pdf
- [23] Woods, D., Dekker, S., Cook, R., Johannesen, L. & Sarter, N. (2010). *Behind human error*. Second Edition. Farnham, Surrey. England: Ashgate Publishing Limited
- [24] Kumar, T., Porambage, P., Ahmad, I., Liyanage, M., Harjula, E. & Ylianttila, M. (2018). Securing Gadget-Free Digital Services. *IEEE Computer*. 66-77. Doi: 10.1109/MC.2018.2876017
- [25] EY (2018). Is cybersecurity more than protection?. EY Global Information Security survey 2018-19. *Reporte*. Recuperado de: [https://www.ey.com/Publication/vwLUAssets/ey-global-information-security-survey-2018-19/\\$FILE/ey-global-information-security-survey-2018-19.pdf](https://www.ey.com/Publication/vwLUAssets/ey-global-information-security-survey-2018-19/$FILE/ey-global-information-security-survey-2018-19.pdf)
- [26] Higuera, J. (2017). La gestión de la ciberseguridad como estrategia: el peligro de la vulnerabilidad y cómo mitigarla. *Harvard Deusto*. 12-19.
- [27] Sabillón, R. & Cano, J. (2019). Auditorías en Ciberseguridad: Un modelo de aplicación general para empresas y naciones. *Revista Ibérica de Sistemas y Tecnologías de Información*. 32. 33-48.
- [28] Roberts, S. & Brown, R. (2017). *Intelligence-Driven Incident Response. Outwitting the Adversary*. Sebastopol, CA. USA: O’Reilly
- [29] Ahmad, Z., Ong, T., Liew, T. & Norhashim, M. (2019). Security monitoring and information security assurance behaviour among employees: An empirical analysis. *Information & Computer Security*. Doi: 10.1108/ICS-10-2017-0073
- [30] Thakur, K., Tseng, J. & Hayajneh, T. (2019). Cyber Security in Social Media: Challenges and the Way Forward. *IEEE IT Professional*. 21(2). 41-49. Doi: 10.1109/MITP.2018.2881373.

Jeimy J. Cano M. Ingeniero y Magíster en Ingeniería de Sistemas y Computación por la Universidad de los Andes. Especialista en Derecho Disciplinario por la Universidad Externado de Colombia. Ph.D in Business Administration por Newport University, CA. USA y Ph.D en Educación por la Universidad Santo Tomás en Colombia. Profesor universitario a nivel de pregrado, especialización, maestría y doctorado en Colombia y varios países de Latinoamérica en temas como computación forense, evidencia digital, gobierno y gestión de la seguridad de la información, gerencia de sistemas de información, criminalidad informática, auditoría de TI, ciberseguridad y protección de datos personales desde 1996 a la fecha. ORCID: <https://orcid.org/0000-0001-6883-3461>

Amenazas Cibernéticas a la Seguridad y Defensa Nacional. Reflexiones y perspectivas en Colombia

M. E. Realpe¹ y J. Cano²

Resumen— Las amenazas cibernéticas impuestas por las tecnologías disruptivas en las Fuerzas de la Ley del Estado, sugieren una peligrosa tendencia hacia la Seguridad y Defensa Nacional. Para mitigar los riesgos en este campo, se requiere una estrategia integral para contrarrestar, si es necesario, la resistencia a los ataques disruptivos y destructivos. En efecto, este documento presenta un análisis general de la situación actual en el campo de la Ciberdefensa en Colombia, donde se identifican las ciberamenazas latentes y emergentes mediante el instrumento denominado Ventana AREM, enfrentando siete (7) tecnologías disruptivas analizadas por los investigadores del MIT a corto plazo, con el fin de proponer una estrategia militar de Ciberdefensa que permita responder a las ciberamenazas con una visión integral, sistémica y prospectiva.

Palabras clave — Ciberdefensa, Tecnologías disruptivas, Ciberamenazas, Estrategia.

I. INTRODUCCIÓN

LA CONVERGENCIA de tecnologías digitales, físicas y biológicas marcan el advenimiento y despliegue de la Cuarta Revolución Industrial o Revolución Digital, generando un cambio veloz y a gran escala en los sistemas informáticos que actualmente se conocen [23]. Este escenario genera un espacio propicio para la incubación de amenazas cibernéticas, que pueden poner en riesgo la prosperidad económica y social de un país, así como su Seguridad y Defensa Nacional.

Dichas amenazas cibernéticas son múltiples y muy variadas, mutan sus vectores de operación en cuestión de segundos, y se soportan en la masificación de las tecnologías disruptivas. Aunado a este comportamiento, los perpetradores de ataques (o adversarios) no descansan en su intento por buscar nuevas y mejores formas de explotación para lograr sus objetivos de manera más fácil y eficiente, creando un escenario complejo y dinámico que debe ser afrontado por las Fuerzas de Ley, de la forma más adecuada, tanto para anticipar como evitar que se produzca una afectación a la Seguridad Nacional. Se debe agregar que, como lo afirma el autor Schwab Klaus, la revolución digital en su esencia, “No cambia lo que hacemos, sino que cambia lo que somos” [22], lo que marca un inicio para repensar y evolucionar el concepto de la Defensa Nacional al nuevo entorno operacional llamado ciberespacio, con el fin de comprender la dinámica de las vulnerabilidades y los retos ante la nueva revolución digital. En este nuevo dominio, es posible que un ataque cibernético realizado a las

plataformas tecnológicas que soportan los servicios esenciales brindados a la población debilite o impida la gobernabilidad de un país, imposibilite la prestación de servicios esenciales ocasionando sufrimiento y, en otros casos, la muerte e incluso desequilibre la economía; hechos que en conjunto pueden desestabilizar la Seguridad y Defensa Nacional. En consecuencia, es necesario considerar el ciberespacio como un nuevo escenario de confrontación bélica, que combinado con los dominios de tierra, mar, aire y espacio constituirán una capacidad y superioridad relevante de cualquier Estado o Nación.

Con respecto a la evolución y el crecimiento tecnológico, es apropiado decir que éstos toman cada vez menos tiempo por lo que, para hablar del futuro del ciberespacio, este documento establece como referencia el año 2022, considerando base el estudio realizado por Instituto Tecnológico de Massachussets (MIT) [23] y las reflexiones de Klaus Schwab (2018), en su libro “La cuarta revolución industrial”, donde se pronostica que “las innovaciones tecnológicas más importantes están a punto de generar un cambio trascendental en todo el mundo y esto es inevitable” [22]. En este sentido, este artículo plantea una estrategia de defensa cibernética a corto plazo para Colombia, de tal manera que, pueda ser adoptada para obtener victorias tempranas y contar con una ventana estratégica de acciones preventivas.

En consecuencia, la Estrategia Militar de Ciberdefensa para las Fuerzas Militares de Colombia que se plantea de cara a las amenazas cibernéticas latentes y emergentes que imponen las tecnologías disruptivas, parte de un análisis de capacidades existentes en las instituciones responsables de la Ciberdefensa Nacional de las Fuerzas Militares, a nivel de doctrina, organización, material, personal e infraestructura u otras variables que permitan determinar la situación actual y así, proyectar un estado deseado. De igual forma, se realiza la identificación y estudio de las amenazas cibernéticas latentes y emergentes que imponen las tecnologías disruptivas a la Ciberdefensa Nacional en Colombia basado en el instrumento llamado la “Ventana de AREM” [13].

Finalmente, con los resultados obtenidos se presenta un esquema general de una propuesta militar para desarrollar capacidades cibernéticas en las Fuerzas Militares, fundamentada en el modelo de planeamiento por capacidades DOMPILEM (Doctrina, Organización, Material, Personal, Infraestructura, Liderazgo, Entrenamiento, Mantenimiento), que permitan configurar unas capacidades que permitan el despliegue operativo necesario para responder ante las nuevas amenazas cibernéticas. Por razones de confidencialidad, Seguridad y Defensa Nacional, este artículo solo presenta un

1. M. E. Realpe. Escuela Superior de Guerra, Colombia. Milena.realpe@esdegue.edu.co

2. J. Cano. Académico y profesor universitario, Colombia. jjcano@yahoo.com

esquema general de fortalecimiento cibernético y algunos contenidos necesarios para su comprensión.

II. CONSIDERACIONES DE LA CIBERDEFENSA NACIONAL

Hoy en día, la tecnología y el internet hacen parte integral en el desarrollo económico, político y social de Colombia y por ende de su Seguridad y Defensa Nacional. Es por esto, que la Ciberdefensa Nacional, debe ser asumida con un enfoque sistémico y multidimensional que requiere un trabajo conjunto y coordinado con los diferentes sectores e instituciones del país para dar cumplimiento a los objetivos e intereses del Estado.

En este sentido, sus Fuerzas Militares, requieren abordar el ciberespacio como un ámbito estratégico, operativo y táctico, para organizar, entrenar y equipar a sus hombres, con el fin de aplicar medidas de prevención, disuasión, contención, protección y reacción, que permitan fortalecer las capacidades de Ciberdefensa, para enfrentar las amenazas o ataques cibernéticos que puedan afectar la infraestructura crítica cibernética del país y poner en riesgo la seguridad nacional, la defensa de la soberanía y el orden constitucional del Estado, así como causar daños masivos, debilitar la economía, y/o dañar la moral pública y la confianza.

Para tal fin, el Ministro de Defensa Nacional de Colombia del año 2012, aprobó la creación y activación del Comando Conjunto Cibernético-CCOCI [21] con la función principal de ejercer la Ciberdefensa de la Nación y conducir operaciones militares cibernéticas a nivel estratégico, para ser garantes de la Seguridad y Defensa de la Nación en el ciberespacio. De igual forma, se ordenó la creación de estructuras organizacionales al interior de cada Fuerza denominadas Unidades Cibernéticas, así: una en el Ejército Nacional, otra en la Armada Nacional y una tercera en la Fuerza Aérea Colombiana, con las cuales el CCOCI ejecutará y coordinará actividades de Ciberseguridad y operaciones de Ciberdefensa del país.

En la actualidad, Colombia cuenta con una estructura organizacional en materia de Ciberdefensa en los niveles estratégico, operacional y táctico. De acuerdo con la primera edición del Manual de Ciberdefensa Conjunta del CCOCI [4], la estructura organizacional de Ciberdefensa se encuentra bajo la responsabilidad del Comando General de las Fuerzas Militares (COGFM) a través del Comando Conjunto Cibernético (CCOCI) que depende de la Subjefatura de Estado Mayor Conjunto Operacional (SEMCO) y de las Unidades Cibernéticas del Ejército Nacional, la Armada Nacional y la Fuerza Aérea Colombiana, de aquí en adelante las Fuerzas.

En efecto, el CCOCI fue activado en octubre de 2012 como ente rector para realizar el direccionamiento, planeación, coordinación, integración, ejecución y sincronización de Operaciones Cibernéticas. Dicho Comando, tiene la misión de ejercer la Ciberdefensa y conducir Operaciones Militares

Cibernéticas a nivel estratégico en el ciberespacio. Para tal fin, dicho Comando mantiene relaciones de coordinación con las Unidades Cibernéticas de las Fuerzas, con el propósito de consolidar esfuerzos e integrar capacidades para las Operaciones de Ciberseguridad y Ciberdefensa [18]. En el Ejército Nacional se creó el Grupo de Apoyo de Comunicaciones y Ciberdefensa, en la Armada Nacional, se creó la Dirección Cibernética Naval, en la Fuerza Aérea Colombiana se creó la Dirección Cibernética Aérea; así mismo de manera coordinada, conjunta y combinada realizan operaciones para la Defensa y Seguridad Nacional.

Adicionalmente, estas Unidades Militares lideran trabajos interinstitucionales en materia de Infraestructuras Críticas Cibernéticas, con el propósito de avanzar en el desarrollo de Políticas y Planes de Protección y Ciberseguridad para sus activos estratégicos nacionales. Como resultado, se han logrado establecer lazos de confianza y amistad entre los diferentes sectores del país que permitieron conformar redes de colaboración para compartir información de amenazas y alertas tempranas como medida preventiva para evitar la materialización de amenazas o ataques sobre las infraestructuras cibernéticas del país. Sin embargo, pese a los esfuerzos realizados, se evidencia la carencia de unas políticas y directrices integrales, positivas y comunes que describan y detallen las oportunidades y desafíos en el quinto dominio de la guerra, el ciberespacio, para empoderar a cada sector en materia cibernética, estableciendo objetivos estratégicos y líneas de acción claras que permitan configurar un control institucional en materia cibernética para conseguir los objetivos nacionales.

III. AMENAZAS CIBERNÉTICAS DE CARA A LAS TECNOLOGÍAS DISRUPTIVAS

La actualidad tecnológica exige un mundo de sistemas complejos e interconectados que traen como consecuencia el crecimiento de las amenazas cibernéticas a ritmos exponenciales y a velocidades desconcertantes. En este contexto, la Ciberdefensa toma un papel preponderante que involucra diversos actores, dándole el eje fundamental a las Fuerzas Militares, para proponer una visión que asegure la defensa cibernética nacional, así como la resiliencia y continuidad de las Infraestructuras Críticas Cibernéticas, en el quinto dominio de la guerra: el ciberespacio [5]. Por su parte, la convergencia tecnológica, la densidad digital, los productos y servicios digitalmente modificados, entre otros, han incrementado los niveles de riesgo cibernético nacional. En este contexto, se revela que estamos frente a una serie de retos que exigen altos niveles de dependencia tecnológica, que despliega un escenario de oportunidades, pero también complejo y desafiante de cara al crecimiento exponencial de las amenazas y vulnerabilidades cibernéticas a la Ciberdefensa Nacional [12].

En este sentido, Colombia está adoptando tecnologías de información y tecnologías de operación para optimizar el desarrollo de sus procesos productivos, económicos,

sociales, políticos e incluso de Seguridad y Defensa Nacional. Dichos procesos se encuentran frente a un escenario de cambios disruptivos, en los cuales se hace necesario adaptarse y prepararse para anticipar y evitar (en la medida de lo posible) daños de alto impacto en las Infraestructuras Críticas Cibernéticas del país, que afecten directamente la prestación de los servicios básicos a la población y, por consiguiente, comprometan la prosperidad económica y social de la Nación. Es una realidad que un gran número de fuerzas de disrupción e innovación están marcando el futuro inmediato, pero también dan lugar al surgimiento de nuevas amenazas y ataques en el ciberespacio que podrían llegar a afectar infraestructuras cibernéticas provocando incluso daños físicos. Por lo tanto, los perpetradores de ataques no descansan en su intento por buscar nuevas y mejores formas de explotación para lograr sus objetivos de manera más fácil, eficiente y anónima.

Para iniciar a hablar de tecnologías disruptivas es necesario empezar por definir las: “Las tecnologías disruptivas son “tecnologías cuya aplicación rompe con los patrones que existía hasta el momento”, apunta Carlos Martín, profesor del *Master of Science in IT Strategic Management* de la UPF Barcelona School of Management. Por su parte, Clayton M. Christensen, profesor de Harvard Business School, acuñó el término “tecnología disruptiva”. En su libro de 1997, *The Innovator's Dilemma*, Christensen, separa la nueva tecnología en dos categorías: sostenida y disruptiva. La tecnología sostenida se basa en mejoras incrementales a una tecnología ya establecida. La tecnología disruptiva carece de refinamiento, a menudo tiene problemas de rendimiento porque es nueva, atrae a un público limitado y puede que aún no tenga una aplicación práctica probada (tal fue el caso de la "máquina de habla eléctrica" de Alexander Graham Bell, que ahora llamamos el teléfono).

Dicho de otra manera, el concepto de “Tecnologías Disruptivas”, implica, entre otras cosas una revolución tecnológica, un cambio en la manera de hacer las cosas. Un estudio realizado por el Instituto tecnológico de Massachusetss (MIT), define siete tecnologías que cambiarán el mundo, las cuales serán objeto de estudio del presente documento: [23], así:

Machine Learning (aprendizaje automático)

De acuerdo con el estudio realizado por el MIT, existen 3 enfoques diferentes: a) Perspectiva de la computadora, programas que "aprenden", b) Una perspectiva relacional, abarca reconocimiento de patrones basado en computadora, modelado estadístico y análisis para la toma de decisiones, y c) Perspectiva holística, combina algoritmos informáticos, patrones estadísticos e inteligencia artificial. Adicionalmente, hay tres principios tecnológicos que, cuando se agrupan, desbloquean los aspectos clave de cada perspectiva: computación en la nube, grandes datos y analítica e inteligencia artificial. Cuando se analiza el

Machine Learning desde esta perspectiva se determina inofensiva y favorable, sin embargo, el futuro parece muy distinto cuando se analiza desde la Seguridad y Defensa Nacional. Es por esto, que se hace necesario cuestionar ¿Qué ocurriría si esta tecnología se despliega para llevar a cabo ataques cibernéticos contra las infraestructuras críticas nacionales?

Nanotechnology (Nanotecnología):

Es una ciencia de la ingeniería que está diseñando y fabricando circuitos y dispositivos extraordinariamente pequeños que se construyen a nivel molecular de materia, típicamente de 1 a 100 nanómetros. Las combinaciones de nanomateriales pueden marcar el comienzo de una nueva era que nos proporciona computadoras y otros dispositivos con un poder de procesamiento hasta ahora inalcanzable. Con la nanotecnología, todo tipo de medios y materiales que no han tenido capacidades de procesamiento y entrega de información pueden convertirse en nuevos portales de comercio y comunicación. Los materiales de nanoingeniería adoptarán nuevas propiedades físicas que son muy diferentes y más útiles que las propiedades físicas de sus estados naturales. Esta tecnología, también se desarrolla en el límite de lo desconocido, y como tal se sujeta a la Seguridad y Defensa al umbral de la incertidumbre. Al manipular los materiales a esta escala, los riesgos pueden ser impredecibles e imperceptibles a los sentidos, incrementando los grados de complejidad en lo que refiere a afectaciones éticas, legales, ambientales y de salud entre otras para Colombia.

Robótica

La robótica no es una tecnología nueva en sí mismo, pero en la última década, la robótica ha sufrido una transformación radical impulsada por tres características: [23]

- Precisión: La capacidad de realizar tareas extremadamente exactas.
- Agilidad: La capacidad de realizar una variedad de tareas de forma rápida y fácil.
- Inteligencia: La capacidad de adquirir y aplicar nuevos conocimientos y habilidades.

Los robots son cada vez más utilizados en todos los sectores, en un sin número de tareas en educación, agricultura, salud, comercio, industria, turismo e incluso en la Seguridad y Defensa. Su rápida evolución pronto hará que la colaboración entre seres humanos y máquinas sea una realidad cotidiana que incremente la eficiencia, precisión y velocidad y reduzca los riesgos de tareas peligrosas minimizando los costos. Las personas que abogan por el desarrollo y despliegue de los sistemas de armamento autónomos normalmente resaltan varias ventajas militares. En primer lugar, los sistemas de armamento autónomos actúan como multiplicadores de Fuerza. Es decir, se necesita un menor número de soldados para cumplir una misión dada

y se incrementa la eficacia de cada soldado. En segundo lugar, los defensores adscriben la expansión del campo de batalla a los sistemas de armamento autónomos, que permite que el combate alcance áreas que previamente eran inaccesibles. En tercer lugar, los sistemas de armamento autónomos pueden reducir el número de bajas al quitar a los combatientes humanos de las misiones peligrosas [18]. Cuando se analiza esta tecnología desde el polo opuesto a la beneficiosa es atinado pensar que las afectaciones para la Seguridad y Defensa de una Nación podrían ser demoledoras. ¿En quién podría recaer la responsabilidad cuando se despliegan los sistemas de armas que generan daños a la población, o cuando un sistema sea capaz de asumir su propio control?

Impresión en tres dimensiones:

Consiste en crear un objeto físico mediante la impresión capa por capa de un modelo o un dibujo en 3D. Con el tiempo las impresoras 3D superarán los obstáculos de velocidad, costo y tamaño y su uso será más generalizado. Esta tecnología ofrece ventajas y también varias desventajas, relacionadas a la creación propia de productos de forma automatizada mediante impresoras 3D. Entre las ventajas están la flexibilidad y prototipado rápido, reducción de costos, personalización, nueva industria, así como aplicaciones múltiples aún por descubrir. Para analizar las desventajas es preciso acoger una óptica desde la orilla de los malintencionados, donde existe la posibilidad de crear objetos tales como armas de fuego, drogas o elementos de alta peligrosidad que deja entrever el peligro de llegar a una proliferación de elementos y una falsificación sin límites de elementos que puedan ser usados en contra de las Fuerzas de Ley.

Biotechnología

Es el uso de sistemas vivos y organismos para desarrollar o fabricar productos. En la actualidad, los avances en tecnología digital, ingeniería genética, informática, tecnología celular y ciencias químicas están ampliando enormemente los límites de la biotecnología. La noción de la ingeniería de las células vivas y el surgimiento de la industria de las ciencias de la vida cambiará radicalmente los límites de la atención médica, la agricultura y los productos químicos. La biotecnología apunta a ser una de las áreas de mayor importancia en la sociedad, particularmente en el siglo XXI. Pese a las ventajas que aporta esta tecnología a la humanidad, también trae consigo una serie de riesgos, principalmente en el medio ambiente y en la salud como la modificación de la estructura del genoma humano, clonación y la manipulación del material genético de nuestra especie entre otros.

Wireless mesh networks (Redes Inalámbricas)

Las redes de malla inalámbricas (WMN) son circuitos ad hoc de conectividad inalámbrica en los que solo un

dispositivo requiere una conexión a Internet. Estas son redes inteligentes de dispositivos inalámbricos que se pueden formar, dispersar y reformar según el comando del usuario. Las WMN se crean de abajo hacia arriba mediante conexiones entre dispositivos, sus capacidades de auto formación y auto curación aseguran una comunicación robusta y confiable en cualquier lugar a bajo costo y sin infraestructura fija. Las WMN amplían la informática generalizada integrada en el IoT (Internet of Things – Internet de las Cosas) haciéndolo más dinámico. Esta tecnología ofrece múltiples beneficios, entre los más significativos están: la innovación y menor costo, al tiempo que los contras se consolidan en que no existe sistemas de regulación para controlar este tipo de tecnología y con ello se integran un sin número de delitos cibernéticos.

Pervasive Computing (Computación Omnipresente)

También conocida como computación ubicua, la cual brinda información, medios, contexto y poder de procesamiento, sin importar la ubicación. Esta clase de tecnologías se caracteriza por amplias redes de microprocesadores conectados o incrustados en objetos cotidianos, los datos se integran y se intercambian en las redes públicas. La computación omnipresente es la tecnología que impulsa Internet de las cosas (IoT), pero es más preciso pensar en ella como el motor de todo el internet. Las capacidades de información, intercambio y colaboración de estas redes no se limitan a ningún dispositivo o ubicación fija; se distribuyen por todo el mundo. Además, el factor de forma de la informática dominante puede ser móvil, usable o implantable. El IoT como toda tecnología disruptiva, conlleva una serie de ventajas y de desventajas. Entre los riesgos más significativos que aporta esta tecnología a la Seguridad Nacional están: la falta de compatibilidad, la complejidad que representan y la falta de privacidad y seguridad, considerando que existe la posibilidad de que el software pueda ser vulnerado y los datos personales mal empleados.

En consecuencia, cada una de estas tecnologías marcarán un cambio profundo y sistémico en la sociedad, facilitarán los procesos y harán más eficiente la vida de los colombianos. Sin embargo, también traerán consigo novedosas amenazas y peligrosos ataques de tipo cibernético que pondrán en riesgo a la Nación.

IV. LA VENTANA DE AREM Y EL ANÁLISIS DE LAS AMENAZAS CIBERNÉTICAS

En esta sección, se analizaron las amenazas cibernéticas haciendo uso del instrumento denominado “Ventana de AREM” [2], dado que, a diferencia de los marcos tradicionales de riesgos que se basan en riesgos y amenazas conocidas, ofrece una visión más amplia de éstos y las vulnerabilidades cibernéticas, con el fin de avanzar hacia una gestión y gobierno de los riesgos empresariales basado más en las posibilidades, que en las probabilidades.

A pesar de que el autor plantea en su instrumento cuatro tipos de amenazas o riesgos así: Conocidos, focalizados, latentes y emergentes, para el caso que nos ocupa en el presente estudio se limitará al análisis de las amenazas y riesgos latentes y emergentes que presentan las tecnologías disruptivas para Colombia a corto plazo, a fin de plantear una estrategia militar que permita dar respuesta y ser efectiva de cara a un análisis prospectivo a corto y mediano plazo.

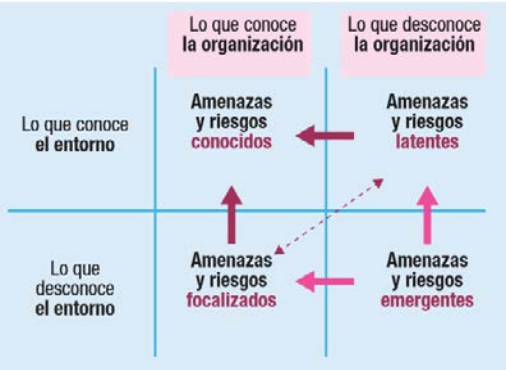


Figura 1. La Ventana de AREM [13].

En primer lugar, partiendo de las definiciones de amenaza latente (el analista se ha enterado que tal amenaza existe y no sabe si la organización tiene alguna estrategia de mitigación) y amenaza emergente (el analista nunca había escuchado de tal amenaza) [13], para aplicar este instrumento un analista de riesgos recolectó y analizó durante un tiempo (generalmente un año), material suficiente de tendencias visibles y emergentes sobre nuevas y posibles amenazas o vectores de ataque, las cuales se valoraron y priorizaron de acuerdo con su nivel de impacto en la organización, novedad del ataque y capacidad de respuesta actual de las Unidades de Ciberdefensa. Con base en este resultado, se estableció una lista de posibles amenazas y riesgos cibernéticos contrastadas con cada una de las siete tecnologías disruptivas mencionadas en la sección anterior, a fin de identificar las posibles amenazas que podrían afectar la Seguridad y Defensa Nacional en un futuro inmediato, las cuales fueron valoradas por los investigadores asistidos con la Ventana de AREM. Posteriormente se priorizaron basado en su impacto y compilaron los diferentes análisis, identificando las mayores coincidencias en las ubicaciones en los diferentes cuadrantes, para luego afinarse con el conocimiento del autor, en las realidades concretas del país y sus niveles de aseguramiento actuales.

El resultado, es una vista sistémica de los riesgos claves a tener en cuenta (dos cuadrantes), enriquecida con la vista de los investigadores y los analistas de riesgos y amenazas en el contexto de la Ciberdefensa Nacional, donde se obtuvo los resultados definidos en la tabla I:

TABLA I AMENAZAS LATENTES Y EMERGENTES	
Amenaza Latente	Amenaza Emergente
Ciberguerra	Guerra Autónoma
LAWS Lethal Autonomous Weapons (Robots Militares, embarcaciones de superficie y submarinas, drones autónomos, sistemas satelitales autónomos)	Ciberarmas de destrucción masiva
Sistemas Autónomos	

Riesgos Latentes

La Ciberguerra: es sin duda, el extremo más grave del espectro de los problemas de seguridad planteados en el ciberespacio donde los actores son los Estados. La ciberguerra puede permitir a los actores lograr sus objetivos políticos y estratégicos en menos tiempo, con mínimos riesgos y a bajo costo. La guerra cibernética o ciberconflicto se configura en la actualidad, como una posibilidad de intervención político-estratégica más eficiente, en la que habrá más posibilidades de que los daños provocados sean menores en comparación con las armas convencionales. Además de eso, será también más difícil de identificar/controlar el origen de los ataques. Por tantas ventajas, las actividades que se suceden en el ciberespacio atraen el interés de los Estados, con el objetivo de utilizarlas como instrumentos de política exterior. El conflicto sigue siendo una extensión de la voluntad política y ahora tiene un nuevo dominio en el que sus formas pueden participar. Colombia, no se encuentra ajena a esta amenaza.

Las armas letales autónomas: (LAWS, por sus siglas en inglés, Lethal Autonomous Weapons). La ONU define como arma autónoma una herramienta capaz de “localizar, seleccionar y eliminar objetivos humanos sin intervención humana”. Estas, son el resultado de la aplicar la inteligencia artificial a la búsqueda de soluciones en el ámbito del enfrentamiento militar; del combate. Las verdaderas armas autónomas letales o LAW, se enfrentarán al enemigo sin órdenes humanas y decidirán qué hacer para evitar los obstáculos que pretendan impedir la ejecución de su misión. Por ejemplo, un pequeño tanque podría patrullar las calles de una ciudad y actuar sin un piloto que decidiera por él. Estas opciones aun no existen, pero ya han desatado una gran polémica y un debate internacional. Hay LAW marinas o acuáticas (embarcaciones de superficie y submarinas), aéreas (drones autónomos que seleccionan por sí mismos los objetivos) y terrestres (algunos incluso con forma humana). El riesgo radica cuando sea la máquina mediante procedimientos autónomos la que elija los objetivos a abatir.

Sistemas Autónomos: Un reconocido científico y académico de la Universidad de California, llamado Stuart Russell, explica que los sistemas autónomos a su juicio es el problema actual de la inteligencia artificial. Ahora presente en un sinnúmero de dispositivos que podrían ser manipulados

para infiltrar sistemas informáticos de otros Estados, empresas o industrias nacionales críticos para la sociedad sin intervención humana. La posible existencia de robots, capaces de disparar las 24 horas del día de manera indiscriminada: es uno de los escenarios que han dejado de pertenecer exclusivamente a las películas y series de ciencia ficción y se ha convertido en una preocupación recurrente de organismos internacionales como la ONU o la Convención de Ciertas Armas Convencionales. En los próximos años cada arma tendrá la capacidad de decidir autónomamente a quién matar: uno de los grandes dilemas éticos de la inteligencia artificial aplicada a la guerra. [9]

Riesgos Emergentes:

La Guerra Autónoma, es el uso de la inteligencia artificial en la guerra. Dicho de otra forma, una guerra autónoma acaba por transformarse en una guerra de recursos donde el humano no interfiere directamente. Simplemente es la interacción de Sistemas Autónomos entre sí, en este orden de ideas se hace necesario establecer pautas para la innovación, que impidan emprender un viaje tecnológico sin retorno.

Las Ciberarmas de Destrucción Masiva, si bien actualmente solo nueve Estados (supuestamente) poseen armas nucleares, las armas cibernéticas pueden ser obtenidas, desarrolladas o utilizadas por cualquier estado o actor no estatal; son relativamente baratos, seguros y fáciles de operar. Esto tiene dos consecuencias. Primero, las armas cibernéticas pueden convertirse en un nuevo tipo de Arma de Destrucción en Masa, o tal vez sería mejor llamarlas Armas de Destrucción Masiva. Es de esperar que dentro de unos años, gracias a la rápida y continua digitalización del mundo, los ciberataques puedan dañar a sociedades enteras. Es posible que las armas cibernéticas no puedan causar el mismo nivel de destrucción mortal que las armas nucleares, pero pueden ser muy eficaces. Piense en un sabotaje combinado y serio de la energía y los suministros de agua, así como en las comunicaciones [28].

Finalmente, y con base en los resultados obtenidos se propone contar con una Estrategia Militar basada en objetivos que consideran como fundamento el desarrollo de capacidades y la aplicación de medidas de contención y mitigación, los cuales sirvieron de base para la formulación de los objetivos estratégicos de la Estrategia propuesta.

V. FORTALECIMIENTO DE LA CIBERDEFENSA NACIONAL

La Estrategia Militar de Ciberdefensa asegura un enfoque conjunto, integral y sistémico que permita integrar los diferentes frentes y aristas relacionadas con la Ciberseguridad y Ciberdefensa Nacional para mejorar la eficacia operacional en el ciberespacio, proporcionando principios fundamentales que guíen el empleo de las Unidades de las Fuerzas Militares de Colombia (FF.MM.) hacia un objetivo común. Es necesario aclarar que las Fuerzas Militares de Colombia es el término

usado para denotar colectivamente todos los componentes del Ejército, Armada y Fuerza Aérea [4].

Se resalta entonces, que a pesar de que la moderna tecnología ha revolucionado la mayor parte de las dimensiones materiales de la guerra desde el siglo XIX, la lógica de los conflictos permanece básicamente inalterable. Esto explica porque obras como “De la Guerra” de Carl Von Clausewitz [16] y “El Arte de la Guerra” de Sun Tzu [25] permanecen como relevantes marcos conceptuales para el estudio de la política y la estrategia incluso en nuestros días. Por lo tanto, a pesar de que la Estrategia Militar es una disciplina en constante evolución, la estrategia propuesta no puede ni debe ignorar las enseñanzas de los clásicos y debe contemplar los principios de la guerra a los que se acogen las Fuerzas Militares de Colombia, los cuales se encuentran enunciados en el Manual Fundamental de Referencia del Ejército MFRE 3-0 enfocadas para su empleo en el ámbito conjunto. La aplicación de estos principios clásicos en la conducción de Operaciones Conjuntas se define y amplía en el Manual de Operaciones Conjuntas. Adicionalmente se considera el modelo de guerra tradicional, donde los Estados-Nación luchan entre sí por razones tan variadas como la gama completa de sus intereses nacionales.

Las Operaciones Militares en la guerra tradicional normalmente se centran en las Fuerzas Militares de un adversario para influir finalmente en el Gobierno de éste. Con el caso cada vez más raro de guerra formalmente declarada, la guerra tradicional típicamente involucra operaciones militares Fuerza-a-Fuerza, en las cuales los adversarios emplean una variedad de fuerzas convencionales y fuerzas de operaciones especiales entre sí en todos los dominios físicos (incluyendo el ciberespacio). Un dominio es el territorio sobre el cual se impone o se ejerce el control (terrestre, marítimo, aéreo, espacial y el ciberespacio) [7]. Bajo este enfoque, y con base en los niveles de la guerra², se definió y clarificó la relación entre los objetivos nacionales, el enfoque operacional y las tareas tácticas, considerando que no hay demarcaciones o límites definidos entre estos niveles, pero que ayudan a los comandantes a diseñar y sincronizar operaciones, asignar recursos y tareas al comando apropiado. En otras palabras, el propósito estratégico, operacional o táctico del empleo depende de la naturaleza del objetivo, la misión o la tarea [17].

Es por esto, que la Estrategia de Ciberdefensa propuesta se centra en la construcción y fortalecimiento de capacidades para el desarrollo de Operaciones Cibernéticas para operar en un dominio donde se disputa el acceso al ciberespacio o donde podría ser imperativo afectar el acceso a las comunicaciones contra cualquier adversario que intenta hacer daño a los intereses nacionales en tiempos de paz, crisis o conflicto para impedir los ataques y hacer la defensa del país, dando cumplimiento a las funciones asignadas al CCOCI

² Niveles de la Guerra: a) Estratégico nacional, b) Teatro estratégico, c) Operacional y d) Táctico

mediante el CONPES 3701 “Lineamientos de Política para Ciberseguridad y Ciberdefensa del año 2011[10].

Mediante esta estrategia se propone planificar, desarrollar y utilizar las capacidades cibernéticas de las Fuerzas Militares de manera efectiva, y asegurar que las Operaciones en el ciberespacio se produzcan de manera consistente con los valores que promueve Colombia dentro de los límites que establece a la Fuerzas Militares, la legalidad y legitimidad a nivel nacional como internacional. En pocas palabras, se proporcionará dirección, coherencia y lineamientos con un enfoque integral para el desarrollo de capacidades militares en el ciberespacio en los próximos años. Para este fin se plantearon cinco (5) objetivos estratégicos, así:

Objetivo Estratégico 1 (OE1): Fortalecer el pie de Fuerza requerido para dar respuesta a las amenazas cibernéticas.

Objetivo Estratégico 2 (OE2): Defender las Infraestructuras Críticas Cibernéticas Nacionales de la actividad cibernética maliciosa.

Objetivo Estratégico 3 (OE3): Desarrollar capacidades cibernéticas para mejorar las ventajas militares en Colombia.

Objetivo Estratégico 4 (OE4): Conformar Ligas de Ciberdefensa con los sectores público y privado.

Objetivo Estratégico 5 (OE5): generar Alianzas Estratégicas a nivel nacional e internacional.

Estos objetivos estratégicos se analizaron contrastados con los componentes de capacidad abreviados en la sigla DOMPILEM (que será detallada más adelante). Estos componentes son aquellos aspectos que deben observarse en el planeamiento por capacidades, tanto en las soluciones materiales como en las no materiales, ante los requerimientos de la Fuerza para afrontar los retos presentes y futuros. Esta sigla también es empleada en la OTAN con una “I” adicional al final para indicar interoperabilidad, o con una “P” para incluir aquellas políticas gubernamentales o institucionales que afectan la capacidad [7].

La doctrina es generalmente el primer aspecto, ya que a menudo es el más fácil y rápido de actualizar y puede afectar drásticamente el desarrollo de las operaciones. En algunos casos, el impacto de los cambios en los otros componentes no puede realizarse plenamente sin alterar de manera significativa la doctrina. Además, la doctrina también puede servir como base para la evolución en las otras categorías DOMPILEM, con el fin de establecer los planes y tareas necesarias para alcanzar las metas: [7]

Doctrina: cómo se combate.

Organización: cómo es el diseño de la fuerza.

Material: elementos necesarios para equipar las fuerzas con el fin de que puedan operar de manera efectiva.

Personal: recurso humano necesario para combatir en la guerra, enfrentar contingencias o participar en operaciones de paz.

Instalaciones: bienes inmuebles.

Liderazgo y educación: cómo preparar a los comandantes en cada uno de los escalones para conducir el combate a través del desarrollo profesional.

Entrenamiento: cómo debe ser la preparación para el combate desde la formación básica hasta la formación individual de especialistas y el entrenamiento en los diferentes escalones.

Mantenimiento: actividades que se requieren para el sostenimiento de la capacidad en el tiempo.

Con el propósito de identificar con claridad los componentes necesarios para alcanzar los objetivos estratégicos propuestos y con esto garantizar el cumplimiento de la Estrategia Militar de Ciberdefensa, se hizo necesario realizar un análisis DOFA (Debilidades, Oportunidades, Fortalezas, amenazas) que permitiera tener una visión global de la situación real de las Unidades de Ciberdefensa de las Fuerzas Militares y consecuentemente con los resultados obtenidos definir las necesidades para cada uno de los componentes.

Posteriormente, y con base en los requerimientos más significativos analizados en cada matriz, se definió la matriz principal que refleja los componentes más significativos en cada uno de los objetivos estratégicos, así:

TABLA II
RESUMEN ANALISIS OBJETIVOS VS DOMPILEM

OBJ.	D Doctrina	O Organización	M Material	P Personal	I Instalaciones	L Liderazgo	E Entrenamiento	M Mantenimiento
OE1	Manuales conjuntos y por Fuerza	Un nuevo componente militar	Hardware Software	Militar y Civil	Oficinas Laboratorios Salas	Estrategias Expertos Especialistas Técnicos	Perseutante	Plataformas Cibernéticas
OE2	Planes de Protección	Equipos interdisciplinarios	Plataformas compartidas	Militar y Civil	Propias y de terceros	Expertos en activos estratégicos sectoriales	Sistemas Control Industrial TI TO	Sistemas de Defensa Cibernética
OE3	Manual Operaciones Cibernéticas	Equipos de Batalla	Equipo de combate cibernético	Comandos Cibernéticos	Sala Mando y Control cibernético	Plan de Carrera Cibernética	Simulación Olimpiadas Cibernéticas	Plataformas de Operaciones Cibernéticas
OE4	Manual de Apoyo Cibernético	Conformación Ligas	Especializado Alto Nivel	Sector Público y Privado	Salas de Intercambio Cibernético	Expertos Cibernéticos	Entrenamiento Interinstitucional	Sistemas de Apoyo
OE5	Acuerdos, convenios	Nacionales Internacionales	Especializado Alto Nivel	Nacional e Internacional	Propia y de Terceros	Internacionalistas	Cursos Internacionales	Sistemas de Intercambio

Esta tabla, refleja los elementos clave que permitirán que las Unidades de Ciberdefensa de las Fuerzas Militares, fortalezcan el desarrollo de capacidades para desarrollar operaciones militares cibernéticas efectivas para fortalecer la seguridad y defensa del país en el ciberespacio.

VI. CONCLUSIONES

El ciberespacio es, por naturaleza, no es un espacio seguro y protegido, y, por tanto, es vulnerable a las amenazas o ataques cibernéticos latentes o emergentes, lo que puede resultar en pérdidas significativas para los sectores económico, político, social y/o constituir una seria amenaza para la defensa nacional o los intereses nacionales, por lo tanto el desarrollo de capacidades en el ciberespacio se constituye en una prioridad para la Seguridad y Defensa de Colombia, que cada vez más, es dependiente de la tecnología, al tiempo que el despliegue de operaciones militares en el ciberespacio es una necesidad para el avance de los modelos de defensa actuales.

Para tener éxito y resultados verificables en las misiones cibernéticas que aseguren la Seguridad y Defensa Nacional, se hace necesario que las Fuerzas Militares realicen operaciones conjuntas entre las diferentes Fuerzas, coordinadas con la Policía Nacional, interinstitucionales con organizaciones tanto del sector público como privado y combinado con las naciones amigas, trabajando de la mano con los organismos binacionales y multilaterales, apoyados siempre que se requiera por las Ligas de Ciberdefensa.

La propuesta de la Estrategia Militar de Ciberdefensa es una respuesta efectiva a los riesgos y amenazas a los que se ve enfrentada la Seguridad y Defensa del país de cara a las tecnologías disruptivas, con ello, se planteó un modelo sistémico basado en objetivos estratégicos analizados en cada uno de los componentes del Modelo DOMPILEN. Con esto fue preciso delimitar y definir prospectivamente hacia donde deben ir las Fuerzas Militares a fin de desarrollar capacidades militares para el desarrollo de Operaciones Cibernéticas, soportadas en un marco legal y constitucional.

REFERENCIAS

- [1] Accenture. (2015). *Guiding digital transformation*, 16.
- [2] William C. (s.f.). *Impact of Alleged Russian Cyber Attacks*. Tomado de <https://nsarchive2.gwu.edu/NSAEBB/NSAEBB424/docs/Cyber-027.pdf>
- [3] Clarke, J. and K. R. (2011). *Guerra en la Red* (1 edición).
- [4] COGFM. (2018). *Manual Fundamental Conjunto MFC 1.0*.
- [5] DoD. (2018). *National Cyber USA*, (September).
- [6] DoD Bulgaria. (2010). *White paper on Defence and the armed forces of the Republic of Bulgaria*, 71.
- [7] EJC. (2017). *Manual Fundamental Doctrina*. (Centro de Doctrina del Ejército - CEDOE, Ed.) (Vol. 1). Bogotá: Imprenta Militar del Ejército Restricciones. Tomado de www.cedoe.mil.co
- [8] Vergara GD, E. (2017). *Operaciones militares cibernéticas*. Buenos Aires.
- [9] Guerrero, J. (n.d.). *Armas autónomas, la amenaza fantasma*. Recuperado June 17, 2019, de <https://www.muyinteresante.es/revista-muy/noticias-muy/articulo/armas-autonomas-la-amenaza-fantasma-461487840117>
- [10] Gobierno Colombia (2017). *CONPES 3701*, 91. Recuperado de <https://colaboracion.dnp.gov.co/CDT/Conpes/Economicos/3701.pdf>. Gobierno de Colombia.
- [11] J, R. (2013). *Political Guidance for Cyber Defence* Portugal, 31976–31979.
- [12] Cano, J. (2018). *IT-Insecurity: Pronósticos de seguridad de la información 2018*. Recuperado Feb 1, 2018, de <http://insecurityit.blogspot.com.co/2017/10/pronosticos-de-seguridad-de-la.html>
- [13] Cano, J. (2017) *La ventana de AREM*. Una estrategia para anticipar los riesgos y amenazas en ciberseguridad empresarial. ISACA Journal. 5. Recuperado de: <https://www.isaca.org/Journal/archives/2017/Volume-5/Pages/the-arem-window-spanish.aspx>
- [14] Cano, J. (2018) *Ciberdefensa Empresarial: Un marco conceptual y práctico en un entorno digitalmente inestable*. En García, P., Barragán, R. & Fuentes, N. M. (2018) *Actas XV Reunión Española de Criptología y Seguridad de la Información*. 96-101. ISBN: 978-84-09-02463-6. Recuperado de: https://www.researchgate.net/publication/328191823_Ciberdefensa_empresarial_Un_marco_conceptual_y_practico_en_un_entorno_digitalmente_inestable From: <https://www.researchgate.net/publication/328191823>.
- [15] Aguilar, J. (2017). *Industria 4.0: la cuarta revolución industrial*. (S.A. MARCOMBO, Ed.). Marcombo. Retrieved from <https://www.casadellibro.com/libro-industria-40-la-cuarta-revolucion-industrial/9788426725684/5991036>
- [16] Clausewitz, K. (2010). *De la Guerra*, 239. Recuperado de <http://www.biblioteca.org.ar/libros/153741.pdf>
- [17] MFC1.0 (2018). *Manual Fundamental Conjunto 1.0* Manual Reservado Comando General Fuerzas Militares.
- [18] MCC (2016). *Manual Ciberdefensa Conjunto Fuerzas Militares*. Manual Reservado.
- [19] Marchant, Allenby, Arkin, Barrett, Borenstein, Gaudet, , ... Meara, (2011). *Science and Technology Law*.
- [20] Netherlands Ministry of Defence. (2012). *The Netherlands - The defence cyber strategy*, 20. Recuperado de <http://www.defensie.nl>
- [21] Resolución Ministerial No. 7436 de 2012. Ministerio de Defensa de Colombia Documento Reservado.
- [22] Klaus S. (2018). *The Fourth Industrial Revolution*, by Klaus Schwab | World Economic Forum. Recuperado marzo 11, 2018, de <https://www.weforum.org/about/the-fourth-industrial-revolution-by-klaus-schwab>
- [23] Segars, A. H. (2018). *Seven Technologies Remaking the World*. MIT Sloan Management Review, (March).
- [24] Strategy Department Defence. (2014). *Cyber Security Strategy for Defence*, 1–18. Recuperado de [https://ccdcoe.org/sites/default/files/strategy/Belgian Defence Cyber Security Strategy.pdf](https://ccdcoe.org/sites/default/files/strategy/Belgian%20Defence%20Cyber%20Security%20Strategy.pdf)
- [25] Sun Tzu. (2003). *El Arte de la Guerra*. Retrieved from <http://www.biblioteca.org.ar/libros/656228.pdf>
- [26] The Department of Defense DoD of the United States of America. (2015). *Strategic Cyber Defense*.
- [27] The Department of Defense DoD of the United States of America (2018). *Strategic Cyber Defense*. [2] Achtert, E. et al. Global Correlation Clustering Based on the Hough Transform. *Statistical Analysis and Data Mining*. vol 1(3), pp. 111-127. 2008
- [28] Van Der Meer, S. (2016). *Cyber Warfare and Nuclear Weapons: Game-changing Consequences?*, (Dic), 36–38. Recuperado de <http://www.techworld.com/news/security/cambridge-researchers>.

Milena Elizabeth Realpe Diaz Oficial del Ejército Nacional, de profesión Ingeniera de Sistemas con énfasis en Telecomunicaciones, especialista en Seguridad de Redes de Computadores, en Seguridad Física y de la Informática, así como en Seguridad de la Información. Es Magister en Seguridad de la Información por la Universidad de los Andes y actualmente estudiante de la Maestría en Ciberseguridad y Ciberdefensa de la Escuela Superior de Guerra en Colombia.

Jeimy José Cano Martinez Ph.D, CFE. Profesor universitario a nivel pregrado, especialización, maestría y doctorado en Colombia y varios países de Latinoamérica en temas como computación forense, evidencia digital, gobierno y gestión de la seguridad de la información, gerencia de sistemas de información, criminalidad informática, auditoría de TI, ciberseguridad y protección de datos personales desde 1996 a la fecha.

GESTIÓN DE LA SEGURIDAD Y ANÁLISIS DE RIESGOS

Realizando una Revisión Sistemática de Metodologías ISRA orientadas a la Seguridad TIC. Periodo 2014-2019

L. E. Sánchez, A. Santos-Olmo, V. Figueroa, D.G. Rosado, E. Fernandez-Medina

Resumen – La sociedad de la información depende cada vez más de los Sistemas de Gestión y Análisis del Riesgo al que se encuentran sometidos sus principales activos de información. Poder disponer de estos sistemas es crítico para una correcta protección de los Sistemas de Información de las compañías. Sin embargo, hoy en día desconocemos la situación de las metodologías, modelos y estándares de riesgo TIC, así como las carencias que tienen los mismos. De esta forma, las empresas no saben cómo seleccionar el modelo de Análisis de Riesgos TIC más adecuado para su compañía. En este artículo, presentamos los resultados obtenidos tras aplicar el método de investigación “Revisión Sistemática de la Literatura” de las propuestas científicas orientadas a las llamadas Metodologías ISRA (Information Security Risk Analysis) publicadas en los últimos 5 años.

Palabras clave — Cybersecurity, Information Systems Security Management, ISRA, Information Security Risk Analysis, ISMS, Risk Analysis.

I. INTRODUCCIÓN

Hoy en día, los elementos digitales o las infraestructuras (computadoras, redes, contenidos, etc.), son elementos cada vez más complejos y dependientes de las TI que están en el centro de nuestras vidas y constituyen los pilares esenciales de nuestras infraestructuras de comunicación, económicas, sociales e institucionales. La seguridad y la mitigación de amenazas dentro de esos sistemas se han convertido implícitamente en una parte fundamental para el ciudadano (para preservar su privacidad), para la empresa (para proteger los activos y transacciones digitales) y para los estados (para proteger sus infraestructuras críticas y asegurar la continuidad del gobierno y servicios gubernamentales, etc.) [1-3], y en especial en ciertos sectores como el de la salud [4], o el de la educación [5].

Para proteger estos sistemas recurrimos a la gestión de la seguridad, que según [6] puede definirse como un sistema de gestión usado para establecer y mantener un entorno seguro de la información. El objetivo principal de un SGSI (Sistema de Gestión de Seguridad de la Información) es afrontar la puesta en práctica y el mantenimiento de los procesos y procedimientos necesarios para manejar la seguridad de las tecnologías de la información. Estas acciones incluyen la identificación de las necesidades de seguridad de la información y la puesta en práctica de estrategias para satisfacer

estas necesidades, medir los resultados y mejorar las estrategias de protección.

La definición de un SGSI es una tarea ardua y compleja que requiere un proceso previo de definición en la compañía donde se quiere establecer. Una de las fases más importantes para la implantación de un SGSI es la de Análisis y gestión del riesgo, que para algunos investigadores es una de las fases más críticas [7].

Todas las organizaciones que utilizan tecnologías de la información tienen problemas con la seguridad de su sistema de información. El primer paso en el proceso de protección de un sistema de información es la identificación y clasificación de los recursos o activos de información que necesitan protección, porque son vulnerables a las amenazas, y para realizar este paso necesitamos contar con sistemas de análisis y gestión de riesgos adecuados [8]. Diferentes investigadores destacan que la gestión del riesgo es un proceso esencial en cualquier modelo de gestión empresarial [9, 10], y que la información es un activo valioso que se espera que esté protegido [11].

Un análisis de riesgos es un proceso sistemático para estimar la magnitud de los riesgos a los que está expuesta una organización, para saber qué decisión tomar ante una posible eventualidad [12]. Para ello, se seleccionan e implementan salvaguardas para poder conocer, prevenir, impedir, reducir o controlar los riesgos identificados. Esto es lo que se entiende como gestión de riesgos.

De forma más técnica, el análisis de riesgos permite determinar cómo es, cuánto vale y cómo de protegidos se encuentran los activos. En coordinación con los objetivos, estrategia y política de la organización, las actividades de gestión de riesgos permiten elaborar un plan de seguridad que, implantado y operado, satisfaga los objetivos propuestos con el nivel de riesgo que acepta la dirección.

Actualmente se están realizando muchas investigaciones sobre análisis de riesgos, y muchas de ellas intentan comparar los métodos clásicos para ver cómo se podrían alinear [13-19]. Otros investigadores han realizado también algunos análisis comparativos de los principales estándares de riesgos con el objetivo de mejorar algunos de sus aspectos -entre ellos podemos destacar [20], o trabajos que relacionan los planes de contingencia con el análisis de riesgos [21]. Uno de los puntos de divergencia entre las metodologías se trata de cómo cuantificar todos estos elementos que forman parte del análisis

L. E. Sánchez, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, Luisenrique@sanchezerespo.org

A. Santos-Olmo, Departamento I+D+i, Sicaman Nuevas Tecnologías y Marisma Shield, Tomelloso (Ciudad Real), España, Asolmo@sicaman-nt.com

V. Figueroa, OPTIC – Gobierno de la Provincia, Neuquen, Argentina, vfigueroa@neuquen.gov.ar

D.G. Rosado, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, David.Grosado@uclm.es

E. Fernandez-Medina, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, Eduardo.FdezMedina@uclm.es

de riesgos. A continuación destacamos algunas problemáticas identificadas por otros investigadores:

- Según Alcántara [22], uno de los problemas que se presentan es que los riesgos de los activos de información tienen una naturaleza compleja. La gestión del riesgo de la seguridad de la información se aborda mediante diferentes enfoques. Existe una importante carencia sobre cómo implementar sistemas que gestionen estos riesgos.
- Para Shamala [23], la seguridad de la información se ha convertido en un elemento esencial para que las organizaciones de todo el mundo eliminen los posibles riesgos en sus sistemas de información mediante la evaluación de riesgos de seguridad de la información (ISRA). Sin embargo, la existencia de numerosos tipos diferentes de métodos de evaluación de riesgos, estándares, pautas y especificaciones disponibles hace que las organizaciones afronten de forma desalentadora la tarea de determinar el método más adecuado para satisfacer sus necesidades.
- Ford [24] destaca que, en base a su experiencia como consultor experto en la materia, la mayoría de las industrias que visita desconocen cómo abordar el tema de la Gestión de riesgos de IT, y la mayoría recurren a realizar un análisis de riesgos obligados por regulaciones como la SOX, HIPAA, etc., y no por el valor diferencial que este tipo de sistemas le puede generar.
- Según Derakhshandeh [25], la seguridad se está convirtiendo cada vez más en un foco crítico en los sistemas de información. Con más redes, movilidad y teletrabajo, existe una mayor necesidad de una evaluación de los riesgos técnicos y de seguridad.
- Dehkhoda [26] analiza la posibilidad de unir métodos tradicionales de análisis de riesgos como IRAM2 con los análisis CBA (Cost-Benefit Analysis), ya que entiende que los unos no pueden funcionar sin los otros.
- Duricu [27], se centra en la necesidad que marca la nueva legislación europea de privacidad de datos (GDPR) de realizar un análisis de riesgos y la necesidad de crear un nuevo modelo, dado que los modelos tradicionales (OCTAVE Allegro, ISO, NIST) no se adaptan a la casuística de esta nueva legislación.
- Para Shamala [28], la seguridad de la información se ha convertido en un punto crítico para las organizaciones de todo el mundo, ya que realizan negocios en un entorno interconectado y rico en información. Por lo tanto, las organizaciones desean eliminar los posibles riesgos en sus organizaciones mediante la evaluación de riesgos de seguridad de la información (ISRA). Los ISRA les permite identificar y priorizar los activos de información y garantizar que se utilicen mecanismos de control efectivos para los activos de información de alta prioridad, pero los métodos actuales de ISRA tienen limitaciones críticas ya que adoptan simplemente una perspectiva técnica. Los métodos ISRA disponibles actualmente funcionan en una vista limitada de los activos de información, y por tanto es necesario desarrollar nuevas taxonomías de activos para ellos.
- Según Wangen [29], gran parte del debate en torno a la gestión de riesgos en seguridad de la información (InfoSec) ha sido a nivel académico, donde la cuestión de cómo los profesionales ven los problemas predominantes es un elemento esencial que a menudo se deja sin explorar. Por lo tanto, este artículo representa una visión inicial de cómo los profesionales de riesgos de InfoSec ven el campo de evaluación de riesgos de Información (ISRA). El estudio presenta los resultados de un estudio de 46 participantes donde han reunido datos sobre problemas conocidos en ISRA. También destacan que la teoría de los “cisnes negros” (eventos catastróficos poco probables) no suele ser tenida en cuenta dentro de estos sistemas. Por último los investigadores determinan que los métodos actuales no son suficientes para resolver los problemas y que son necesarios nuevos métodos más avanzados.
- Para Haythorn [30], la realización de una evaluación de riesgos es un paso esencial para las organizaciones a fin de garantizar que existan controles adecuados para proteger los activos que son críticos para las funciones comerciales. La evaluación de riesgos puede ser una tarea muy compleja, que requiere múltiples metodologías y recursos para realizar análisis cuantitativos y cualitativos basados en evidencia fáctica y opinión subjetiva. En última instancia, la organización tiene la responsabilidad del análisis preciso y las medidas de control. La necesidad de una evaluación de riesgos precisa ha creado múltiples marcos de referencia que las organizaciones pueden utilizar para cubrir sus necesidades. Es responsabilidad de los profesionales de seguridad de la información dentro de la organización analizar múltiples marcos y utilizar los métodos que sean ideales para cada caso.
- Según Pandey [31], cualquier activo de información, cuando está conectado al mundo exterior, es vulnerable a los ataques. Los ataques son causados principalmente por amenazas que tienen el potencial de explotar vulnerabilidades. Cualquier tipo de daño a estos activos causa riesgos y es uno de los factores más importantes para la organización. El riesgo de ataques maliciosos a la seguridad del software ha aumentado considerablemente y es muy necesario evitarlo. La máxima “antes es mejor” se ha convertido en el orden del día. Por lo tanto, este estudio se realizó en vista de la importancia de la evaluación de riesgos en la fase de requisitos de SDLC (Software Development Life Cycle).
- Rea-Guaman [32] destaca que, en ciberseguridad, la identificación de riesgos es una parte fundamental porque esta actividad no es exclusiva de la ciberseguridad y es difícil saber cuáles son los riesgos en esta área concreta. Este estudio tiene como objetivo identificar si existen taxonomías de riesgo en ciberseguridad.
- López [33] se centra en el Riesgo Dinámico y destaca que la aplicación de procesos de Análisis y Gestión de Riesgos en el ámbito de los Sistemas de Información es una práctica común que permite la planificación en un

momento puntual de tiempo de las acciones preventivas frente al riesgo a corto, medio o largo plazo, pero con un considerable potencial actualmente desaprovechado para facilitar la toma de decisiones en tiempo real frente a eventos o incidentes de seguridad. Este trabajo hace un recorrido por las principales corrientes que buscan sacar partido a este potencial, englobadas principalmente bajo el concepto de Análisis de Riesgos Dinámico.

- Para Ganin [34], los evaluadores y gerentes de riesgos enfrentan muchos desafíos difíciles relacionados con los nuevos sistemas cibernéticos. Entre estos desafíos se encuentran la naturaleza en constante cambio de los sistemas cibernéticos causado por los avances técnicos, su distribución a través de los dominios físicos, de información y sociocognitivos, y las complejas estructuras de red que a menudo incluyen miles de nodos.
- Según Smojver [35], numerosos métodos existentes de gestión de riesgos de seguridad de la información (ISRM) difieren mucho en su enfoque, complejidad de uso, nivel de detalle y aplicabilidad a organizaciones de diferentes tamaños y modelos de negocio. La selección de un método que se ajuste a los requisitos de una organización puede ser un proceso complejo e intensivo en recursos, con una posibilidad significativa de decisión subóptima.
- Beckers [36] destaca la importancia de constatar que los estándares de seguridad y de gestión de riesgos puede ser un desafío, en parte porque las descripciones de lo que se debe realizar suelen ser genéricas y deben ser perfeccionadas por expertos en seguridad. Eliminar esta ambigüedad requiere mucho tiempo para los expertos en seguridad, ya que tienen que interpretar todas las tareas requeridas en el estándar por su cuenta.
- Para Shedden [37] existen muchas metodologías para evaluar los riesgos de seguridad asociados con fugas no autorizadas, modificación e interrupción de información en una organización determinada. Argumentan que la orientación tradicional de estas metodologías, hacia la identificación y evaluación de los activos de información técnica, oscurece los riesgos clave asociados con el cultivo y despliegue del conocimiento organizacional. Basándose en la literatura de gestión del conocimiento, el estudio sugiere mecanismos para incorporar estas consideraciones basadas en el conocimiento en el alcance de las metodologías de riesgo de seguridad de la información.
- Según Rot [38], el riesgo relacionado con la aplicación de las tecnologías de la información en los negocios crece junto con el aumento de la correlación empresarial con sus clientes, socios comerciales y operaciones subcontratadas. El progreso tecnológico genera dependencias que evocan el crecimiento de la diversidad, la complejidad, la falta de descripción y la cantidad de factores de riesgo. En inversiones insuficientes en seguridad de la información, el tema de la gestión de riesgos de TI se vuelve más importante,

concentrándose en buscar la proporción óptima entre las amenazas y los costes de la protección de los sistemas de TI. En un desarrollo tan dinámico de las Tecnologías de la Información, el tiempo necesario para una reacción apropiada ante el riesgo se acorta de forma determinante. La falta de una preparación adecuada puede llevar a la empresa al colapso, por lo que la reacción adecuada al riesgo constituye la posibilidad de supervivencia y desarrollo de la empresa. El problema de la gestión de riesgos de TI es un problema muy complejo. Una de las etapas más importantes de este proceso es el análisis de riesgos, utilizado para la optimización y la minimización de las pérdidas relacionadas con el riesgo.

- Saripalli [39] plantea la necesidad de modernizar los modelos de riesgos para que se puedan adaptar a entornos como el Cloud Computing [39]
- Por último Li y Sicari [40, 41], proponen la necesidad de avanzar en el análisis de riesgos para IoT, tanto desde el punto de vista de modernizar los modelos existentes como la necesidad de añadir sistemas inteligentes de valoración del riesgo, tales como las redes neuronales.

Toda esta información y cómo se lleva a cabo el proceso está recogido en lo que se denominan metodologías de análisis de riesgos. Aunque es cierto que existe un gran número de metodologías para este tema, se puede decir que la mayoría tienen puntos en común. Según [12] las metodologías de análisis de riesgos tienen como punto de partida identificar formalmente los elementos a proteger o aquellos que tienen un valor para la organización, lo que se llamarán activos.

Después de analizar estas propuestas, vemos que es de interés el llevar a cabo una revisión sistemática (RS) de la literatura existente en relación con las Metodologías y Estándares de Análisis de riesgos que se están utilizando actualmente, su estado y evolución histórica, con el objetivo de entender su situación actual.

Como se trata de una revisión sistemática, se sintetiza el trabajo existente de forma que sea coherente [42-44]. En contraste con el proceso habitual de una revisión de la literatura, que se lleva a cabo de manera no sistemática cada vez que alguien acomete una parte particular de una investigación, una RS se desarrolla, como el término denota, de una manera formal y sistemática [45]. Esto significa que el proceso de investigación de una revisión de tipo sistemático sigue una secuencia de pasos metodológicos muy bien definida y estricta, conforme a un protocolo desarrollado apriorísticamente. Ésta se lleva a cabo en torno a un tema central, que representa el núcleo de la investigación, y que se expresa mediante el uso de una pregunta específica, previamente definida, centrado y estructurada. Los pasos metodológicos, las estrategias para recuperar la evidencia y el enfoque en la cuestión se definen explícitamente, de manera que otros profesionales pueden reproducir el mismo protocolo y también pueden juzgar la idoneidad de los estándares elegidos para el caso en cuestión.

Esta revisión sistemática se va a llevar a cabo mediante el uso de las directrices para las revisiones sistemáticas propuestas

por Kitchenham [42-44], que se han mostrado apropiadas para investigaciones de ingeniería del software. Usaremos también un modelo de protocolo de revisión desarrollado por Biolchini et al. [45], que facilita la planificación y ejecución de las revisiones sistemáticas en la ingeniería del software.

El resto del artículo se estructura de la siguiente forma: En la Sección 2 definiremos la pregunta de investigación. En la Sección 3 se explica el método de revisión, que se basa en el protocolo de investigación, y es aquí donde se definirá la estrategia de búsqueda y la selección de estudios. En la Sección 4 definiremos los datos a extraer y que se presentarán en el resumen de los estudios relevantes. En Sección 5 presentaremos los resultados de la revisión y un análisis de los mismos. Finalmente, en la última sección describiremos las principales conclusiones obtenidas.

II. PLANIFICACIÓN DE LA REVISIÓN

En este apartado, se define la pregunta de investigación de forma que se focalice el área de interés del trabajo y queden definidos tanto el problema a tratar como sus principales características.

A. Objeto de la pregunta.

En esta revisión sistemática se pretende localizar trabajos centrados en el desarrollo de modelos y metodologías de análisis de riesgos de carácter general, con el objetivo de entender su estado actual y cuáles son los que se están utilizando actualmente.

B. Premisa de la Revisión Sistemática.

Podemos definir la pregunta de investigación de este trabajo, por tanto, de la siguiente forma:

¿Qué Metodologías y Estándares existen actualmente relacionados con el Análisis de Riesgos y en qué estado se encuentran?

Las palabras y conceptos relacionados que se utilizaron para formular esta pregunta y que fueron utilizados durante la ejecución de la revisión son las siguientes:

“Information Security Risk Assessment”, ISRA
“Information Security Risk Management”, ISRM
Risk management: Risk assessment methodology

En el contexto de la revisión sistemática planificada se van a observar las propuestas existentes sobre modelos y metodologías de análisis de riesgos generales, extrayendo las más importantes y procediendo a un posterior análisis y

comparación de las mismas. La población a analizar se compone de las publicaciones presentes en los repositorios de las fuentes de datos seleccionadas que estén relacionadas con el objetivo de esta revisión.

Los resultados esperados de esta revisión son conocer las propuestas existentes en cuanto a análisis de riesgos generales, para posteriormente analizarlas y conocer qué comparten y en qué difieren, además de identificar necesidades de investigación. Las principales áreas de aplicación que se verán beneficiadas por los resultados de esta revisión sistemática son las relacionadas con la Seguridad de la Información, en especial la Gestión de la Seguridad (concretamente los análisis de riesgos), así como los expertos en seguridad. A tal fin, se proveerá un marco comparativo que permita posicionar adecuadamente las nuevas actividades de investigación en análisis de riesgos.

III. MÉTODO DE LA REVISIÓN

El método de revisión se basa en el protocolo de investigación. En esta etapa definimos la estrategia de búsqueda, qué fuentes se utilizarán para identificar los estudios primarios, si hubo algunas restricciones, cuáles son los criterios de inclusión y exclusión, qué criterios se utilizarán para evaluar la calidad de los estudios primarios y cómo se extraerán y sintetizarán los datos de los estudios.

A. Selección de fuentes.

El objetivo de esta fase es seleccionar las fuentes que se usarán para realizar la ejecución de la búsqueda de estudios primarios.

El criterio para la selección de las fuentes de búsqueda será la posibilidad de consultar los documentos en Internet o en la biblioteca digital de la Universidad de Castilla-La Mancha, que cuenta con libros electrónicos así como con acceso a las bibliotecas digitales de ACM, IEEE, Science@-Direct o Elsevier, entre otros; la inclusión motores de búsqueda que permitan consultas avanzadas y búsqueda por palabras clave; además, editoriales, libros, revistas y conferencias recomendadas por expertos en la materia.

La búsqueda de estudios primarios se llevará a cabo utilizando motores de búsqueda en web, bases de datos electrónicas y búsquedas manuales, tales como búsquedas en una revista/conferencia/libro/publicación específica o en publicaciones de investigación recomendadas por expertos en la materia.

Finalmente, las principales fuentes de lista de fuentes inicial obtenida sobre la cual se ejecutará la revisión sistemática es la siguiente: ACM digital library, IEEE digital library, Science@Direct, Google Scholar, SREIS symposium, ESORICS symposium, REFSQ conference, IEEE International Requirements Engineering Conference, ICSE conference, COMPSAC conference, DEXA conference, WOSIS workshop, ICCSA conference, Requirements Engineering Journal, Computer Standards & Interfaces Journal, Computers & Security.

B. Selección de estudios.

Una vez que se han sido definidas las fuentes, es necesario describir el proceso y el criterio que vamos a seguir en la ejecución de la revisión para la selección y evaluación de los estudios.

En primer lugar, se combinaron las palabras clave seleccionadas con conectores AND y OR para obtener la cadena de búsqueda, como se muestra a continuación:

methodology OR standard OR guidelines
AND
"Information Security Risk Assessment" OR ISRA
AND
"Information Security Risk Management" OR ISRM
AND
"risk analysis" OR "risk management" OR "risk
assessment"

El procedimiento para la selección de estudios empleado comienza con la adaptación de la cadena de búsqueda al motor de búsqueda de la fuente y la ejecución de la consulta, limitando la búsqueda a trabajos publicados en los últimos 5 años (2014 – 2019). Los criterios de inclusión y exclusión deberían basarse en la Pregunta de investigación. El criterio de inclusión actúa sobre los resultados obtenidos al ejecutar la búsqueda sobre la fuente, permitiéndonos realizar una primera selección de documentos que serán considerados en el contexto de la revisión como candidatos a convertirse en estudios primarios. Como criterio de inclusión se realiza principalmente un análisis sobre el título, las palabras claves y el abstract de cada documento, de forma que podemos ver en una primera instancia cómo están relacionadas estas palabras y porqué ha sido seleccionado el estudio. Con este criterio se localizan y eliminan la mayor parte de los resultados obtenidos que no realizan aportaciones sobre análisis de riesgos en el campo de los Sistemas de Información.

El criterio de exclusión actúa sobre el subconjunto de estudios relevantes obtenidos y nos permite obtener el conjunto de estudios primarios. En esta fase nos centramos principalmente en la lectura y análisis del *abstract* del documento y sus conclusiones, teniendo en algunos casos que profundizar en el mismo y realizar una lectura más detallada sobre otras partes del documento, de cara a ver en más detalle de qué trata cada estudio, ver la relación real que presenta con los objetivos buscados y, si es verdaderamente relevante para la revisión, seleccionarlo como estudio primario.

C. Ejecución de la selección.

En este punto, se ejecuta la búsqueda en cada una de las fuentes seleccionadas con el fin de obtener una lista inicial de los estudios para la posterior evaluación aplicando todos los criterios y procedimientos especificados.

Los procedimientos para la selección de los estudios se

aplican a todos los artículos obtenidos a fin de verificar si los estudios se ajustan a los criterios de inclusión y exclusión. Los estudios obtenidos, que corresponden exactamente con todos los criterios de inclusión y exclusión definidos previamente, se detallan en la siguiente sección.

IV. EXTRACCIÓN DE LA INFORMACIÓN.

La información extraída de los estudios debe contener las técnicas, métodos, procesos, medidas, estrategias o cualquier tipo de iniciativa para la adaptación del análisis, gestión o evaluación de riesgos a nivel general.

Los *formularios de información* definidos para esta revisión sistemática contienen la identificación del estudio, la metodología o modelo del estudio, los resultados del estudio, los problemas del estudio y nuestras impresiones generales al respecto.

Algunos de los artículos obtenidos han sido descartados por no ofrecer información relevante, o estar sesgados hacia algunos sectores, aun así consideramos interesantes hacer mención sobre ellos:

- Svatá, V. and M. Fleischmann, "IS/IT Risk Management in banking industry" [46]: En esta investigación el autor analiza la relación de algunos estándares de riesgos con la norma Bancaria de Basilea II para alinearlos en su aplicación al sector financiero.
- Mayer, N., P. Heymans, and R. Matulevicius. "Design of a Modelling Language for Information System Security Risk Management". [47]: En esta investigación el autor plantea la necesidad de diseñar un lenguaje formal para el análisis y gestión de riesgos de información.

A continuación se ofrece una breve reseña de cada uno de los estudios seleccionados mostrados en la sección anterior, de acuerdo con la información extraída obtenida a través de los formularios de información.

1.1. García, F.Y.H. and L.M.L. "Moreta. Maturity Model for the Risk Analysis of Information Assets based on Methodologies MAGERIT, OCTAVE y MEHARI; focused on Shipping Companies" [48].

En la Investigación [48, 49] se propone un nuevo modelo de madurez para el Análisis de Riesgos de los Activos de Información, derivado de las Metodologías MAGERIT, OCTAVE y MEHARI y una orientación sectorial, en concreto a empresas del sector marítimo.

La Investigación propone un modelo de mapa de control compuesto por 11 categorías a evaluar (A. Política de Riesgo, B. Responsabilidad, C. Compromiso de la Alta Dirección; D. Comunicación y Formación; E. Determinación y Valoración de los Activos de Información; F. Identificación y estimación de Amenazas; G. Estimación de Impacto; H. Evaluación del Riesgo; I. Respuesta a los Riesgos; J. Actividades de Control y K. Mejora Continua del Análisis de Riesgo) y 5 Niveles de Madurez (Nivel 1. Inicial; Nivel 2. Repetible; Nivel 3. Definido; Nivel 4. Administrado; Nivel 5. Optimizado).

1.2. Gritzalis, D., et al., “Exiting the Risk Assessment maze: A meta-survey” [50].

En la Investigación se puede ver una comparativa entre 10 metodologías de Análisis de Riesgos, con 3 criterios de comparación.

Se analizan las siguientes metodologías de Análisis de Riesgos: EBIOS, MEHARI, OCTAVE, IT-Grundschutz, MAGERIT, CRAMM, HTRA, NIST SP800, RiskSafe Assessment y CORAS.

Se analizan las siguientes dimensiones: Si incluye las cuatro fases del análisis de riesgos (1. Preparación; 2. Identificación del Riesgo; 3. Análisis del Riesgo; y 4. Evaluación del Riesgo); El modelo de cálculo del riesgo; y el tipo de Análisis que realiza (Cualitativo o Cuantitativo).

El artículo también muestra otras tablas comparativas interesantes entre las metodologías que incluyen su vinculación con otros estándares y costes asociados con el soporte y el software. Finalmente, muestra un análisis de características como: la facilidad de uso, el ciclo de vida de sus versiones, los objetivos respecto al tamaño de las empresas, software que la soporta, adaptabilidad, etc.

1.3. Mrksic Kovacevic, S., “Smart homes from a Risk Management perspective” [51].

En la Investigación podemos ver el trabajo de Tesis realizado sobre una maestría especializada en “Risk Assessment and Management”.

Dentro de esta tesis de maestría, el autor analiza y compara diferentes metodologías y estándares de análisis de riesgos, en particular: OCTAVE Allegro, FAIR, NIST CSF, RaMEX, ISRAM, CORAS y CIRA.

De cada una de ellas se analiza: i) Tipo de metodología; ii) Nivel; iii) Tiempo de implantación; iv) Fortalezas; v) Debilidades; y vi) Perspectivas individuales, empresariales y gubernamentales.

1.4. Wangen, G., C. Hallstensen, and E. Snekenes, “A framework for estimating information security risk assessment method completeness” [52, 53].

En la Investigación podemos ver cómo se plantea la dificultad de comparar los diferentes métodos al no existir criterios unificados de comparación. Por ello se propone la creación de un marco unificado de comparación (Core Unified Risk Framework – CURF), que permitirá comparar diferentes modelos.

Dentro de esta investigación se compararon los siguientes modelos de riesgos: CIRA, CORAS, CRAMM, FAIR, NSMROS, OCTAVE A, ISO27005, NIST 800-30, RISK IT, RAIS y CRDF.

Para cada uno de los modelos se analizan en detalle: i) Los problemas relacionados con la identificación de riesgos; ii) La estimación del riesgo; iii) Problemas relacionados con la Evaluación del riesgo; y iv) Problemas de la Integridad del Riesgo.

1.5. Novoa, H.A. and C.R. Barrera, “Metodologías para el análisis de riesgos en los sgsi” [54].

En la Investigación podemos ver una pequeña comparativa entre algunas de las principales metodologías de Análisis de Riesgos. En concreto, se analiza OCTAVE, MAGERIT, MEHARI, NIST SP 800-30, CORAS, CRAMM y EBIOS, analizando su ámbito de aplicación y las ventajas y desventajas de cada una de ellas, orientándolas a su aplicación en los SGSI, aunque el análisis que se realiza es superficial y no entra a identificar las principales problemáticas que pueden tener estos modelos metodológicos en su aplicación real.

1.6. Santonja Lillo, J., “Análisis y correlación entre probabilidad e impacto de los riesgos” [55].

En la Investigación se realiza una comparativa entre algunas de las principales metodologías de Análisis de Riesgos. En particular, se analizan Octave (las tres versiones), Magerit, Mehari, Cramm y NIST SP 800:30, analizando las ventajas y desventajas de cada una de ellas y las fases que cubren.

Las fases han sido catalogadas en: Caracterización del Sistema, Identificación de Amenazas, Identificación de Vulnerabilidades, Análisis de Controles, Determinación de la Probabilidad, Análisis de Impacto, Determinación del Riesgo, Recomendaciones de Control, Documentación de los resultados, Establecimiento de Parámetros y Necesidades de Seguridad.

1.7. Hashim, N.A., et al., “Risk Assessment Method for Insider Threats in Cyber Security: A Review” [56].

En la Investigación se realiza una comparación entre NIST SP 800-30, OCTAVE, FRAP y CRAMM, mencionando las principales investigaciones que las referencian, así como el tipo de metodología, las fases principales de cada una de ellas y los recursos que requiere su implementación. Dentro de esta investigación, su principal conclusión es que el método que mejor aceptación tiene de los evaluados fue el del NIST por su orientación práctica y su simplicidad, aunque no está exento de deficiencias que deben abordarse.

1.8. Bergvall, J. and L. Svensson, “Risk analysis review” [57].

En la Investigación se realiza una comparación entre Attack trees (método de los investigadores), CRAMM, ISRAM, OCTAVE Allegro, COBRA, Mehari, Magerit y CORAS, clasificándose estas metodologías de Análisis de Riesgos en base a un conjunto de características, que incluyen: Recursos necesarios, número de problemas identificados, tipo de AR (Cuantitativo y Cualitativo) y si tiene soporte para la toma de decisiones (es decir, que establezcan claramente qué acciones son necesarias para mitigar los riesgos).

En las conclusiones de la investigación se destaca que la mayoría de los métodos evaluados carecían de soporte a la decisión, es decir, no especificaban de forma clara las acciones que debían implementarse para mitigar los riesgos, en particular CRAMM, ISRAM, MEHARI y MAGERIT. Esto dio lugar a que cada una de las compañías que formó parte de la investigación (Ericsson, SAAB, Tage Rejmes Bil AB, etc.)

había implementado el análisis de riesgos con métodos diferentes y no unificados.

También se identifica la necesidad de elevados conocimientos para implantar métodos como CRAMM, ISRAM, MEHARI y CORAS.

1.9. Abbass, W., A. Baina, and M. Bellafkih. “Using EBIOS for risk management in critical information infrastructure” [58].

En la Investigación podemos ver la utilización de la metodología EBIOS para la gestión del riesgo dentro de la Infraestructuras Críticas, y dentro del artículo se muestra una comparativa de diferentes metodologías de riesgos. En concreto, se centra en la comparación de OCTAVE, EBIOS, MEHARI, CRAMM y CORAS, analizando características como la fecha de creación, soporte, herramientas, monitorización del riesgo, metodologías de análisis y documentación disponible de las mismas.

La principal conclusión que obtiene es que, a pesar del proceso estructurado, los resultados de los métodos de Análisis de Riesgos son en gran medida informales y no suelen ser suficientemente analíticos. Esta informalidad muestra una falta de automatización, razonamiento y trazabilidad del proceso de Análisis de Riesgos.

1.10. ENISA. “Inventory of Risk Management / Risk Assessment Methods” [59].

En [59] la Agencia Europea para la Ciberseguridad analiza las principales metodologías de Análisis de Riesgos que se utilizan dentro del Marco de la Unión Europea, mostrando sus principales características.

En particular se analizan las siguientes metodologías: Austrian IT Security Handbook, Cramm, Dutch A&K Analysis, Ebios, ISAMM, ISF Methods, ISO/IEC 13335-2, ISO/IEC 17799, ISO/IEC 27001, IT-Grundschutz, Magerit, Marion, Mehari, MIGRA, Octave, RiskSafe Assessment y NIST SP800-30.

Para cada una de ellas se analizan las siguientes dimensiones: Información general, Nivel de referencia del producto, Tarjeta de identidad del producto, Alcance y Características desde el punto de vista de los usuarios.

1.11. Pan, L. and A. Tomlinson, “A systematic review of information security risk assessment” [60].

En [60] se hace una revisión de las publicaciones científicas asociadas al análisis de riesgos, pero centradas en las metodologías NIST SP800-30, ISO27005/ISO27001, OCTAVE e ISRAM. Determina que estos mecanismos de análisis de riesgos tienen importantes carencias, como por ejemplo que no pueden abordar algunos factores importantes, como son la fuga de activos, los activos creados por los usuarios y el conocimiento crítico. También analiza las desventajas de estos modelos a la hora de obtener puntuaciones de riesgo objetivas. Por último plantean la necesidad de investigar los ISRA desde la perspectiva económica, como el análisis de coste-beneficio y la teoría de juegos.

1.12. Shameli-Sendi, A., R. Aghababaei-Barzegar, and M. Cheriet, “Taxonomy of information security risk assessment (ISRA)” [61]

En esta investigación los autores presentan una taxonomía para la evaluación de riesgos de seguridad, construida a partir del análisis de 125 artículos publicados entre 1995 y 2014.

Destacan que uno de los mayores problemas que han encontrado es que, aunque se investiga mucho sobre el riesgo, la mayoría de los modelos propuestos están basados en taxonomías antiguas, dejando a un lado la necesidad de considerar nuevos criterios relacionados con el cambio de tecnologías y del nivel de conocimiento de los atacantes.

Analiza algunas de las principales metodologías de riesgos, como CRAMM, CORAS, OCTAVE, MAGERIT, Microsoft Security Risk Management Guide, MEHARI, ISO27005, NIST SP800-30 y un grupo de proyectos de investigación en riesgos.

Sobre ellos, evalúa la perspectiva con respecto al riesgo, las técnicas utilizadas, valoración, entradas y salidas, valoración de recursos, medición de riesgo y fases de riesgo.

1.13. Ruan, K. – “Introducing cybernomics: A unifying economic framework for measuring cyber risk”. [62]

En esta investigación el autor analiza diferentes modelos de riesgos como ITIL, COBIT, ISO/IEC 27005:2011, ISF (SPRINT y SARA), OCTAVE, NIST 800-53, NIST 800-37, ISO/IEC 31000:2009, COBRA, CORAS, BPIRM, ISRAM, CRAMM, BSI Guide y CORA, y sobre ellos analiza las problemáticas y limitaciones existentes que se resumen en:

- Los métodos actuales se centran en la tecnología y dejan de lado otros factores como las personas, procesos y factores de riesgo socioeconómicos.
- Las estimaciones más precisas a menudo requieren acceso a datos y conocimiento que una sola entidad no posee.
- Los marcos de evaluación de riesgos predominantes como ISO / IEC 27002 están estructurados en función de dominios de control de seguridad, que no son lo suficientemente efectivos para evaluar la preparación de una entidad hacia un conjunto de escenarios de pérdida de alto riesgo desarrollados en torno a activos digitales críticos.
- La proliferación de metodologías de evaluación de riesgos en ausencia de un punto de referencia común ha causado inconsistencias indeseables en la medición del riesgo cibernético.

De esta forma, es necesario plantear otros puntos de vista diferentes para el análisis de riesgos que tenga en cuenta riesgos externos, mayor objetividad, etc.

1.14. Madhavan, K. and R. ManickaChezian, “International Journal of Engineering Sciences & Research Technology a Study on Information Security and Risk Management in IT Organizations”. [63]

El enfoque del presente estudio fue analizar 14 metodologías de análisis de riesgos en detalle y reconocer algunos criterios comunes, con el objetivo de ayudar a las

empresas en la toma de decisiones de cuál puede ser más adecuada para su negocio.

Las metodologías elegidas fueron: IT Grundschutz (BSI 2008), “Standard of Good Practice” (ISF 2005), CRAMM, OCTAVE-O, OCTAVE-A, COBIT, CORAS, ISM3, NIST SP 800-30, ITIL, EBIOS, MEHARI, GAISP y LRAM.

1.15. Radanliev, P., et al., “Economic impact of IoT cyber risk-analysing past and present to predict the future developments in IoT risk analysis and IoT cyber insurance”. [64]

En esta investigación se realiza un análisis de 16 métodos de Análisis de Riesgos y su orientación y validez para los sistemas IoT.

Los métodos analizados fueron: ITIL, COBIT, ISO27005:2011, ISF (SPRINT y SARA), OCTAVE, NIST SP 800-37, NIST SP 800-53, ISO/IEC 31000:2009, COBRA, CORAS, BPIRM, ISRAM, CRAMM, BSI Guide, BS7799 y CORA.

Los resultados de esta investigación conducen a la conclusión de que existen muchos desafíos para comprender los tipos y la naturaleza del riesgo cibernético y sus dependencias / interacciones orientadas a los IoT.

Por otro lado, el estudio concluye que las metodologías clásicas no se adaptan a los riesgos IoT y que es necesario desarrollar nuevas métricas y métodos de valoración de riesgos.

1.16. Acevedo, N. and C. Satizábal. “Risk management and prevention methodologies: a comparison”. [65, 66]

Los investigadores analizan algunas metodologías de gestión y prevención de riesgos (OCTAVE, CORAS, AS/NZS 4360:1999, IS/IEC 27005, CRAMM, MAGERIT, 2 versiones de NIST y BID), realizando una comparación de las etapas que incluyen y determinando si tienen en cuenta el factor humano en el análisis y tratamiento de riesgos.

Entre las conclusiones que obtienen es que solo el 42.85% de las metodologías de gestión de riesgos estudiadas incluyen el factor humano, siendo el NIST la metodología más completa, pero debe completarse con otras metodologías como la BID (Banco Interamericano de Desarrollo - Diagnóstico, prevención y control de la corrupción en los programas de metodología de seguridad cívica).

De las siete metodologías de gestión de riesgos estudiadas, sólo tres consideran el factor humano (Octave, Magerit y la metodología de gestión de riesgos del NIST) y cuatro no lo consideran de manera explícita (CORAS, la metodología del estándar australiano, la NTC-ISO/IEC 27005 y CRAMM).

1.17. Devia, G.A.V. and C.J. Pardo, “Hacia un modelo para la gestión de riesgos de TI en MiPyMEs: MOGRIT”. [67]

La investigación destaca la importancia cada vez mayor del análisis de riesgos, y por ello presenta la armonización de modelos de riesgos de TI (e.g., CRAMM, COBIT, EBIOS, ITIL V3 MAGERIT, OCTAVE, RISK IT) y algunas normas enfocadas en brindar soporte a los riesgos (e.g., ISO/IEC 27000,

ISO/IEC 27005, ISO/IEC 31010, AS/NZS ISO 31000, BS 7799-3:2006, y UNE 71504:2008), realizando un análisis comparativo, de alto y bajo nivel, que permite conocer las características más comunes y representativas de cada uno de ellos.

El análisis realizado permitió evidenciar que la mayoría de las normas y modelos tienen puntos en común, aunque algunas normas presentan procesos más detallados, con un nivel más profundo que otros modelos. Asimismo, se observó que hay normas y modelos con similitudes en la definición de sus procesos, tales como actividades similares entre sí. Por otra parte, también se encontraron algunas actividades que complementaban y mejoraban las descripciones de otras actividades, dando como resultado la característica en la que un modelo es capaz de soportar a otro modelo.

La gestión de riesgos permite evitar el fracaso de proyectos de desarrollo de software, estimulando la terminación del mismo de modo que se incrementa la calidad en los proyectos entregados, reduciendo costos y cumpliendo con las necesidades del cliente, lo que impacta positivamente en su satisfacción. Una buena gestión de riesgos tiene como habilidad entregar a tiempo los productos esperados a partir de las metas que se plantearon y con el cronograma de actividades establecido.

1.18. Alhajri, R.M., et al. “Dynamic Interpretation Approaches for Information Security Risk Assessment”. [68]

El investigador plantea la necesidad de analizar los modelos de riesgos para prever los riesgos probables y llegar a las contramedidas apropiadas, y para ello analiza modelos como OCTAVE, CRAMM, FRAP e ISRAM, además de algunos modelos de investigación.

Las conclusiones que presenta determinan que el factor determinante básico es que el enfoque de evaluación de riesgos más importante es el que incorpora las tres dimensiones de seguridad: confidencialidad, integridad y accesibilidad.

1.19. Korman, M., et al. “Overview of enterprise information needs in information security risk assessment”. [69].

Según los investigadores, los métodos para la evaluación de riesgos en seguridad de la información sugieren a los usuarios recopilar y considerar conjuntos de información de entrada, a menudo notablemente diferentes, tanto en tipo como en tamaño, lo que suele hacer que los análisis de riesgos de las mismas compañías, sobre el mismo alcance y en igual de circunstancias sean diferentes.

Para explorar estas diferencias, este estudio compara doce métodos de análisis de riesgos (IT-Grundschutz, TRA-1, TRITF, CORAS, ISO/IEC27005, MEHARI, TSRMG, MAGERIT, OCTAVE, MG-3, NIST RMF, HMG IA), y se analiza cómo sus sugerencias de entrada se corresponden con los conceptos de ArchiMate, un lenguaje de modelado ampliamente utilizado para la arquitectura empresarial.

Como conclusión, determinan que varios factores podrían explicar las diferencias en las sugerencias de entrada entre los métodos. Por un lado, proporcionar sugerencias detalladas

puede beneficiar a los analistas, ya que identificar información que sea verdaderamente relevante para una evaluación de riesgos es una tarea difícil que merece una visión considerable y un pensamiento amplio. Por otro lado, sugerencias muy detalladas podrían sesgar cognitivamente a los analistas para seguir un esquema establecido que no sea necesariamente completo o equilibrado, lo que podría llevar a pasar por alto elementos de relevancia que de otro modo probablemente se identificarían. Esto último podría pesar especialmente a la luz de los constantes cambios de tecnologías y amenazas. Una alternativa podría ser invitar a los analistas a que identifiquen qué es lo más relevante para el objetivo específico de la evaluación en el día que se ésta realice.

1.20. Fulford, J.E., “What Factors Influence Companies’ Successful Implementations of Technology Risk Management Systems”. [70]

El artículo critica el poco éxito empresarial que tienen las metodologías de análisis de riesgos desarrolladas puramente en el ámbito académico, que no han sido aplicadas en la práctica. Para ello analiza algunas metodologías como OCTAVE, CRAMM, ISO27001, ISRAM, FAIR, CORAS, así como otras metodologías que están en fase de investigación.

Las conclusiones extraídas de las comparaciones de las metodologías y modelos de gestión de riesgos tecnológicos investigados durante la revisión de la literatura incluyeron:

- Algunos métodos de análisis de riesgos (como CORAS y FAIR) utilizan aspectos de diseño de otras metodologías de riesgo y son muy complementarios con las metodologías en uso.
- Los profesionales han desarrollado modelos que típicamente incluyen métodos cuantitativos, pero estos generalmente carecen de un método matemático componente de análisis, como un modelo estadístico, lo que limita la capacidad de esos modelos para determinar las relaciones e impactos de eventos de riesgo operacional de seguridad y tecnología para las operaciones de TI.
- Los modelos profesionales generalmente se construyen para usuarios experimentados con amplios conocimientos de dominio.

1.21. Chen, F., “An Investigation and Evaluation of Risk Assessment Methods in Information systems”. [71]

Los investigadores realizan una revisión de diferentes métodos de análisis de riesgos y de algunas dimensiones, y presentan un marco de trabajo para ayudar a las empresas a elegir el método que más se adapta a sus circunstancias.

Entre los métodos elegidos para el análisis están: OCTAVE, CORAS, CORA, COBRA, Risk Watch, FRAP, COSO ERM y @Risk. Para el análisis posterior, no se tendrá en cuenta ni Risk Watch, ni @Risk, al ser estas herramientas para análisis de riesgos, y no metodologías orientadas al riesgo TIC.

Dentro de las características los investigadores analizan: el objetivo, alcance, flexibilidad, alineamiento con otras normas, precio, tipología, tiempo de implantación, factor humano y usabilidad.

1.22. Pan, L., “Application of a Financial Quantitative Risk Model to Information Security Risk Assessment”. [72]

Los investigadores destacan que muchas organizaciones no pueden identificar sus activos, y mucho menos calcular el riesgo asociado. Esta investigación argumenta que se necesita un nuevo enfoque para la evaluación de riesgos y presenta una alternativa basada en modelos financieros.

Planean también que los estándares de gestión de seguridad de la información son esenciales para la seguridad de la información, ya que permiten contar con un marco completo para asegurar los datos en un nivel apropiado.

Se examinaron cuatro estándares ISRA comúnmente utilizados: OCTAVE, FAIR, ISO 27005 y NIST SP800-30. FAIR presta mayor atención a los métodos de cálculo del análisis de riesgos, mientras que la ISO 27005 prefiere proporcionar un marco completo para la evaluación de riesgos.

Sin embargo, el primer inconveniente de estos ISRA es que son difíciles de ejecutar en periodos cortos, o con seguimientos diario. Son procesos costosos y que requieren rellenar entrevistas y cuestionarios, seguidos por expertos. El segundo problema es que son demasiado genéricos para todos los riesgos de seguridad de la información.

1.23. Nurse, J.R., S. Creese, and D. De Roure, “Security risk assessment in Internet of Things systems”. [73]

Según los investigadores, los sistemas clásicos de protección ya no son adecuados para los entornos IoT, ya que a medida que aumenta la complejidad, la omnipresencia y la automatización de los sistemas tecnológicos, particularmente con el Internet de las cosas (IoT), existe un fuerte argumento sobre la necesidad de nuevos enfoques para evaluar el riesgo y generar confianza. Estos riesgos podrían estar relacionados con los altos grados de conectividad presentes o el acoplamiento de sistemas digitales, ciberfísicos y sociales, por lo que es necesario analizar los modelos existentes y plantear nuevas alternativas que consideren la dinámica y la singularidad de IoT, pero manteniendo el rigor de las mejores prácticas en la evaluación de riesgos.

Para ello analizan métodos como NIST SP800-30, ISO/IEC 27001, OCTAVE y su orientación hacia el IoT.

1.24. Agrawal, V., “A Comparative Study on Information Security Risk Analysis Methods”. [74]

El investigador destaca que el análisis de riesgos es una parte integral de la práctica de gestión y un elemento esencial del buen gobierno corporativo. Hay muchos métodos de análisis de riesgos disponibles en la actualidad, y es una tarea tediosa para una organización (particularmente una pequeña y mediana empresa) elegir el método adecuado. Otro problema, es que aunque hay muchos métodos y herramientas disponibles en este dominio, existen muy pocos inventarios que estén estructurados de acuerdo con un conjunto de propiedades comunes.

Por lo tanto el investigador analiza cuatro métodos de análisis de riesgos y los compara a partir de un conjunto de atributos genéricos como: entrada, resultado, propósito, esfuerzo, escalabilidad, metodología, etc.

Los métodos analizados son: CORAS, CIRA, ISRAM e IS (IS Risk Analysis Based on Business Model).

1.25. Wangen, G.B., “Cyber Security Risk Assessment Practices: Core Unified Risk Framework”. [75].

Los investigadores consideran que las evaluaciones de riesgos permiten reducir la incertidumbre con respecto a eventos futuros con el fin de tomar las mejores decisiones posibles y controlar el riesgo. En la industria, el objetivo es encontrar el equilibrio adecuado en la toma de riesgos en relación con el apetito y la tolerancia al riesgo de la organización. Demasiados controles de seguridad inhibirán la funcionalidad del negocio, y lo contrario conducirá a una exposición inaceptable. Por lo tanto, la investigación aborda varios aspectos de las prácticas de evaluación y gestión de riesgos de seguridad informática y cibernética (ISRA) y contribuye a nuevos problemas de investigación, métodos, modelos y conocimiento dentro de la disciplina.

Entre los desafíos dentro del campo de ISRM destacan la necesidad de elegir entre diferentes métodos ISRA sin una premisa clara, y con una literatura sobre el tema de los problemas en ISRM bastante dispersa.

El estudio encontró que los principales problemas de ISRM según los expertos estaban en la comunicación de riesgos, las medidas de seguridad y el retorno de las inversiones. Mientras que para la evaluación y el análisis de riesgos, encontraron que los problemas clave son la aplicación de métodos cuantitativos y cualitativos, la necesidad de experiencia y la evaluación de activos.

La investigación propone además combinar los métodos cuantitativos (estadísticos) y cualitativos (basado en el conocimiento subjetivo) para poder modelar el ataque y estimar el riesgo. El enfoque se centra en estimaciones cualitativas de activos, vulnerabilidades, amenazas, controles y resultados asociados, junto con un análisis estadístico del riesgo.

Los modelos de riesgos analizados son: CIRA, CORAS, CRAMM, FAIR, NSMROS, OCTAVE-A, ISO27005, NIST SP800-30, Risk IT, RAIS y CRDF.

1.26. MUKAMA, J., “Risk Analysis as a Security Metric for Industrial Control Systems”. [76].

Según los investigadores, a medida que avanza el tiempo y la tecnología, las personas se vuelven más dependientes de los servicios proporcionados por los Sistemas de Control Industrial (ICS). Utilizados principalmente en las industrias de infraestructura crítica, los ICS han realizado y habilitado cientos de servicios esenciales para las personas, el público y las organizaciones a diario.

Para mitigar los riesgos que pueden surgir debido a las vulnerabilidades introducidas en el sistema, los investigadores realizaron una comprensión más profunda de los diferentes ICS, revisaron una serie de enfoques de análisis de riesgos existentes y los categorizamos en términos de su objetivo general, si son cualitativos o enfoques cuantitativos, las etapas de la gestión de riesgos abordados y el alcance en términos de los problemas que abordaron.

Finalmente concluyeron que no existe ningún método de riesgos que sea totalmente adecuado a los ICS, pero que se pueden utilizar NIST y CORAS como enfoques subyacentes para desarrollar un Marco de Análisis de Riesgo Modificado para sistemas ICS (MRAF-ICS). Este marco asigna pesos a todos los activos del sistema para enfatizar la importancia / criticidad del activo en el sistema general. Utiliza el enfoque de modelado de amenazas, FMEA y HAZOP para identificar exhaustivamente las amenazas, los peligros y las vulnerabilidades en el sistema.

Para la investigación se tuvieron en cuenta los siguientes métodos: CRAMM, CORAS, OCTAVE, MEHARI, CSMRA, FMEA&FMECA, HAZOP, HMRM-CI, NIST SP800-30 y ARMS.

1.27. Oppliger, R., G. Pernul, and S. Katsikas, “New Frontiers: Assessing and Managing Security Risks”. [77].

Como hemos visto, existen múltiples investigaciones que destacan que el análisis cuantitativo de riesgos, como se requiere para la evaluación y la gestión de riesgos, funciona mejor en la teoría que en la práctica, y que se necesitan algunos enfoques alternativos [78].

Basado en ese problema se planteó este artículo, y otros asociados, que permitió a los autores realizar un estudio sobre la gestión de riesgos en general, así como sobre el valor intrínseco de la evaluación de riesgos. La mayoría de los investigadores que participaron estuvieron de acuerdo con la hipótesis inicial: los enfoques actualmente implementados para la evaluación de riesgos no funcionan en la práctica y son difíciles o imposibles de aplicar en el campo. Hay una serie de razones para esto, que se pueden ver en el artículo de la IEEE Security & Privacy [78] y que se pueden resumir en que el uso de la teoría de probabilidad y las estadísticas en un campo en constante cambio como la ciberseguridad no tiene sentido.

También analiza la comparación de métodos de Wangen [13] sobre OCTAVE, ISO/IEC 27005:2011 y NSMROS. Oppliger sostiene que la elección de un método sobre otro influye enormemente en el proceso de evaluación resultante.

Otra investigación analizada en el artículo es la de Burnap [79], que plantea el problema de que los métodos actuales de análisis de riesgos están planteados para sistemas independientes, pero los sistemas actuales son interdependientes y complejos. Lo que se requiere para manejar con éxito tales sistemas desde una perspectiva de evaluación de riesgos actualmente no se comprende bien.

Finalmente, analiza las investigaciones de Rossebo [80, 81] el cual presenta una nueva propuesta de análisis de riesgos diseñada para sistemas IoT, que tiene una aplicabilidad mucho más amplia y que puede motivar más investigaciones en este área. Al establecer un conjunto de requisitos para comparar los métodos de evaluación de riesgos existentes para el sector energético, los autores pudieron evaluar los métodos existentes, lo que les permitió identificar la necesidad de un cuarto método, la Metodología de Gestión de Riesgos SEGRID (SRMM), que podría proporcionar un marco de gestión de riesgos para sistema IoT.

V. ANÁLISIS DE RESULTADOS.

En esta sección analizaremos los resultados obtenidos y catalogaremos las características de cada una de las metodologías, estándares y guías asociadas al Análisis de Riesgos que hemos encontrado.

En total se han analizado 27 artículos científicos del periodo 2014-2019, que han permitido identificar 40 métodos, guías y estándares clásicos asociados al Análisis de Riesgos TIC y a nuevos aspectos específicos como el IoT, Smart Grid o ICS, que podríamos considerar la evolución futura de este tipo de sistemas.

En la Tabla 1 podemos ver los 40 modelos identificados durante la revisión sistemática, junto con la organización, país de origen y el número de artículos donde ha aparecido.

Modelo Análisis de Riesgos	Origen	Organización	País	Nº Artículos Investigados Relacionados
ARMS (Automated Risk Management System)	Gubernamental	Defence R&D Canada – Ottawa	Canadá	1 [76]
AS/NZS 4360 / ISO31000:2009 (incluye las guías HB)	Gubernamental	Council of Standards Australia	Australia y Nueva Zelanda	4 [64-67]
Austrian IT Security Handbook (Manual austriaco de seguridad informática)	Gubernamental	Bundeskansleramt (Cancillería federal austriaca)	Austria	1 [59]
BPIRM (Business Process: Information Risk Management)	Empresarial	KPMG	Reino Unido	2 [62, 64]
CIRA (Conflicting Incentives Risk Analysis)	Universidad	Desarrollo como Tesis Doctoral por Lisa Rajbhandari, bajo el nombre de “Conflicting Incentives” as an Alternative Notion of Risk.	Noruega	5 [51-53, 74, 75]
COBRA (The Consultative, Objective and Bi-functional Risk Analysis)	Empresarial	C & A Systems Security Ltd.	Reino Unido	4 [57, 62, 64, 71]
CORA (Cost-Of-Risk Analysis)	Empresarial	International Security Security, Ltd.	Estados Unidos	3 [62, 64, 71]
CORAS (Construct a platform for Risk Analysis of Security Critical Systems)	Empresarial	C&A Systems Security LTD. Information Society Technologies (IST) Programme (Commission of the European	Reino Unido	19 [50-54, 57, 58, 61-66, 69-71, 74-76]

		Communities, Directorate-General Information Society)			
COSO ERM (Enterprise Risk Management — Integrated Framework)	Organización	Committee of Sponsoring Organizations of the Treadway Commission	Estados Unidos	1	[71]
CRAMM (Central computer and Telecommunication Agency Risk Analysis and Management Method)	Gubernamental	CCTA británica (Agencia Central de Comunicaciones y Telecomunicaciones) Insight Consulting	Reino Unido	19	[50, 52-59, 61-67, 70, 75, 76]
Dutch A&K Analysis (Análisis holandés A&K)	Gubernamental	Ministerio holandés de asuntos internos	Países Bajos	1	[59]
EBIOS (Expression of Needs and Identification of Security Objectives)	Gubernamental	ANSSI (Agence nationale de la sécurité des systèmes d'information) DCSSI (Dirección Central de Seguridad de los Sistemas de información)	Francia	6	[50, 54, 58, 59, 63, 67]
FAIR (Factor Analysis of Information Risk)	Empresarial	FAIR Institute	Estados Unidos	6	[51-53, 70, 72, 75]
FRAP (Facilitated Risk Assessment Process)	Universitario	Peltier and Associates	Estados Unidos	3	[56, 68, 71]
GAISP (Generally Accepted Information Security Principles)	Organización	Information Systems Security Association International (ISSA)	Estados Unidos	1	[63]
HMG-IA (HMG IA Standard No. 1 Technical Risk Assessment)	Gubernamental	National Technical Authority for Information Assurance	Reino Unido	1	[69]
IS RA on BM (IS Risk Analysis Based On Business Model)	Gubernamental	Korea Advances Institute of Science and Technology	Corea	1	[74]
ISAMM (Information Security Assessment and Monitoring Method)	Empresarial	Telindus N.V	Bélgica	1	[59]
ISF Method: “Standard of good Practice”, IRAM, IRAM2, SARA, SPRINT, FIRM	Organización	ISF (Information Security Forum) ISF asociación internacional de más de 260 empresas líderes y organizaciones del sector público	Reino Unido	4	[59, 62-64]
ISO/IEC Baselines (incluye ISO/IEC	Organización	ISO	Internacional	16	[52, 53, 59-62, 64-67,

27001 - BS7799-2:2002, ISO/IEC: 27002 - ISO/IEC 17799:2005, ISO/IEC 27005 - ISO/IEC 13335-2/3, BS 7799-3:2006)			(organización con sede en Suiza)		69, 70, 72, 73, 75, 77]
ISRAM (Information Security Risk Analysis Method)	Universidad	National Research Institute of Electronics & Cryptology & TurkeybGebze Institute of Technology	Turquía	8	[51, 57, 60, 62, 64, 68, 70, 74]
IT-Grundschtz (IT Baseline Protection Manual)/BSI Standard 100-2-3	Gubernamental	Oficina Federal de Seguridad de la Información (BSI)	Alemania	6	[50, 59, 62-64, 69]
ITIL (The IT Infraestructura Library)	Organización	Office of Government Commerce (OGC)	Reino Unido	4	[62-64, 67]
LRAM (Livermore Risk Analysis Methodology)	Universidad	Lawrence Livermore National Laboratory (University of California)	Estados Unidos	1	[63]
MAGERIT (Methodology for Information Systems Risk Analysis and Management)	Gubernamental	Ministerio de Administraciones Públicas (Ministerio de Administraciones Públicas de España)	España	12	[48-50, 54, 55, 57, 59, 61, 65-67, 69]
MARION (Metodología de análisis de información y optimización de datos por Niveau)	Gubernamental	CLUSIF	Francia	1	[59]
MCRDF (Microsoft Cloud Risk Decision Framework)	Empresarial	Microsoft	Estados Unidos	3	[52, 53, 75]
MEHARI (Me'thode Harmonise'e d'Analyse de Risques— Harmonised Risk Analysis Method)	Organización	CLUSIF (Francia) transmitido por CLUSIQ (Canadá)	Francia	10	[49, 50, 54, 55, 57-59, 61, 63, 69]
MG-2 (A Guide to Security Risk Management for Information Technology Systems) & MG-3 (A Guide to Risk Assessment and Safeguard Selection for Information Technology Systems)	Organización	CSE (Communications Security Establishment)	Canadá	1	[69]
Microsoft's Security Risk Management Guide	Empresarial	Microsoft	Estados Unidos	2	[61, 69]
MIGRA (Metodología)	Empresarial	AMTC / Elsag Datamat S.p.A	Italia	1	[59]

Integrata per la Gestione del Rischio Aziendale) - (anteriormente se denominaba Defender)					
NIST Family: NIST CSF / NIST RMF / NIST SP800-30 / NIST SP800-37 / NIST SP800-39 / NIST SP800-53	Organización	Instituto Nacional de Estándares y Tecnología (NIST)	Estados Unidos	20	[50-56, 59-66, 69, 72, 73, 75, 76]
NSMROS (Norwegian National Security Authority Risk and Vulnerability Assessment)	Gubernamental	Norwegian Security Act	Noruega	4	[52, 53, 75, 77]
OCTAVE (The Operationally Critical Threat, Asset, and Vulnerability EvaluationSM)/ Octave-S / Octave-A (Allegro)	Universidad	Universidad Carnegie Mellon, SEI (Instituto de Ingeniería de Software)	Estados Unidos	29	[48-73, 75-77]
RAIS (Risk Assessment of Information Systems)	Gubernamental	Norwegian Data Protection Authority's (Datatilsynet)	Noruega	3	[52, 53, 75]
RaMEX (Risk Analysis and Management expert system)	Universidad	University of Birmingham	Reino Unido	1	[51]
Risk IT - ISACA / COBIT Risk (Control Objectives for Information and Related Technology)	Empresarial	ISACA	Estados Unidos	8	[52, 53, 62-64, 67, 69, 75]
RiskSafe Assessment	Empresarial	Platinum Squared Ltd	Reino Unido	2	[50, 59]
SRMM (SEGRID Risk Management Methodology) - SEGRID (Security for Smart Electricity GRIDs)	Gubernamental	ETSI (The European Telecommunications Standards Institute)	UE	1	[77]
TRA (Threat and Risk Assessment Methodology) & HTRA (Harmonized Threat and Risk Assessment Methodology)	Gubernamental	Communications Security Establishment Canada (CSEC) and the Royal Canadian Mounted Police (RCMP)	Canadá	2	[50, 69]

Tabla 1. Modelos identificados durante la revisión sistemática

A continuación, vamos a extraer de forma resumida las principales necesidades y problemas identificados durante la revisión sistemática:

- Controles: Los controles deben formar parte del Análisis de Riesgo, y no sólo de la Gestión del Riesgo: Gran parte de las metodologías consideran que los controles son ajenos al análisis de riesgos, y no lo consideran hasta la fase de gestión de riesgos [75].
- Capacidad Sectorial: Importancia de tener la capacidad en las metodologías de adaptarse a sectores específicos [48].
- Estructuras comunes de riesgo que soporten diferentes metodologías: Ante la diversidad de marcos de trabajo, ser capaces de crear una estructura común que pueda unificar las diferentes metodologías de riesgos [71].
- Los sistemas de riesgos son el núcleo de los sistemas de Gestión de Seguridad [72].
- La simplicidad y la orientación práctica es importante para las empresas [56-58, 72, 75, 77].
- Los métodos deben tener mecanismos de soporte a la toma de decisiones [57].
- Los resultados del análisis de riesgos son informales y poco analíticos, obteniendo puntuaciones de riesgo subjetivas [58, 60].
- Necesidad de contar con orientaciones y perspectivas económicas del análisis de riesgos [60].
- Necesidad de taxonomías de riesgos actualizadas y adecuadas a las nuevas tecnologías (ICSs, IoT, Smart Grids, ...) [61, 64, 73, 76].
- Se dejan de lado perspectivas como personas, procesos y factores de riesgo socioeconómicos [61, 65, 66].
- Necesidad de contar con escenarios de riesgos [61].
- Herramientas: Necesidad de contar con herramientas que faciliten el cumplimiento de las metodologías [71].
- Necesidad de poder contar con mecanismos de selección de metodologías de análisis de riesgos según la compañía [69].
- Necesidad de contar con métricas adecuadas [72].
- Necesidad de que las metodologías académicas se validen en entornos reales [70].
- Necesidad de metodologías adaptadas a las PYMES [74].
- Necesidad de catálogos de elementos, estructurados y que puedan compartirse entre las diferentes metodologías [74].

VI. CONCLUSIONES.

En este artículo se han analizado diferentes metodologías, procesos y estándares de análisis de riesgos TIC (ISRA), extrayendo más de 40 modelos que suelen referenciar y analizar los investigadores actualmente.

De los modelos encontrados, se han identificados unas 20 carencias de relevancia que destacan los investigadores sobre los modelos actuales.

Estas carencias se complementan con otros aspectos relevantes que hemos encontrado al analizar los propios modelos de riesgo, entre las que podemos destacar:

- Sorprende comprobar que metodologías muy referenciadas y recomendadas como CRAMM ya no tienen páginas web accesibles, y que gran parte de la

información facilitada por ENISA sobre las metodologías de análisis de riesgos está ya obsoleta.

- Existen muchos modelos de los analizados que no han sido actualizados en los últimos 10 años, y aun así se siguen utilizando.
- Se han identificado tímidas propuestas para modernizar los análisis de riesgos ante los nuevos retos como el Cloud, IoT, ICSs, etc. Sin embargo, hasta ahora han tenido poco recorrido.
- La mayoría de los modelos encontrados tienen una complejidad alta o media de implantación, no considerándose adecuados para PYMES.
- La mayoría de los modelos se orientan a cubrir las tres dimensiones base (confidencialidad, integridad y disponibilidad), dejando de lado el resto de criterios.

Podemos concluir por lo tanto, que ahora mismo existen importantes carencias dentro de los modelos de Análisis de Riesgos existentes que deben ser afrontadas, desarrollando nuevas metodologías que permitan adaptarse a las circunstancias cambiantes de las TIC.

AGRADECIMIENTOS

Esta investigación ha sido co-financiada por los proyectos *GENESIS - Security Government of Big Data and Cyber Physics Systems ((SBPLY/17/180501/000202)* financiado por el "JCCM- Consejería de Educación, Cultura y Deportes, y Fondos FEDER", del proyecto *ECLIPSE – Enhancing Data Quality and Security for Improving Business Processes and Strategic Decisions in Cyber Physical Systems (RTI2018-094283-B-C31)* financiado por la "Ministerio Economía, Industria y Competitividad y fondos FEDER", y ha contado con el apoyo de las empresas Marisma Shield S.L (www.emarisma.com) y Sicaman Nuevas Tecnologías (www.sicaman-nt.com).

REFERENCIAS

- [1] Le Grand, G. and E. Adar. *White cyber knight—a Risk Assessment tool for network resilience evaluation. in the proceedings of the International Workshop on Complex Network and Infrastructure Protection (CNIP'06), Rome. 2006.*
- [2] Vivas, T., A. Zambrano, and M. Huerta. *Mechanisms of security based on digital certificates applied in a telemedicine network. in 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2008. IEEE.*
- [3] Huerta, M., et al. *Implementation of a open source security software platform in a telemedicine network. in Proceedings of the 9th WSEAS international conference on Advances in e-activities, information security and privacy. 2010. World Scientific and Engineering Academy and Society (WSEAS).*
- [4] Pirrone, J. and M. Huerta. *Security Mechanism for Medical Record Exchange Using Hippocratic Protocol. in World Congress on Medical Physics and Biomedical Engineering 2018. 2019. Springer.*
- [5] Huerta, M., et al. *Design of a building security system in a university campus using RFID technology. in 2017 IEEE 37th Central America and Panama Convention (CONCAPAN XXXVII). 2017.*
- [6] Eloff, J. and M. Eloff, *Information Security Management - A New Paradigm. Annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology SAICSIT'03, 2003: p. 130-136.*

- [7] Steve, E., *An Introduction to information systems risk management*. SANS Institute InfoSec Reading Room. 16: p. 2011.
- [8] Bača, M. and F. Varaždin, *The risk assessment of information system security*. Fakultet organizacije i informatike, Sveučilište u Zagrebu. < dostupno na http://cuc.carnet.hr/cuc2004/program/radovi/a5_baca/a5_full.pdf>, [očitano 07.10. 2010], 2004.
- [9] Restrepo, L.O. and F.J.V. Duque, *Gestión de riesgos en eTOM. Un análisis comparativo con los estándares de riesgo corporativo*. Revista Logos, Ciencia & Tecnología, 2017. 9(1): p. 85-99.
- [10] Ortiz Restrepo, L., V. Duque, and F. Javier, *Gestión de riesgos en eTOM. Un análisis comparativo con los estándares de riesgo corporativo*. Revista Logos Ciencia & Tecnología, 2017. 9(1): p. 85-99.
- [11] Chen, T.M., *Information security and risk management*, in *Encyclopedia of Multimedia Technology and Networking, Second Edition* 2009, IGI Global. p. 668-674.
- [12] Magerit V2, *Methodology for Information Systems Risk Analysis and Management (MAGERIT version 2)*, 2005, Ministerio de Administraciones Públicas (Spain).
- [13] Wangen, G., *Information security risk assessment: a method comparison*. Computer, 2017. 50(4): p. 52-61.
- [14] Fakrane, C. and B. Regragui, *Interactions and Comparison of It Risk Analysis Methods*. in *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech)*. 2018.
- [15] Jeannot, F., *Méthodologies d'évaluation et gestion de risques en sécurité*. Montréal, Canada, Mai 2018, R518, v1.0, 2018.
- [16] Benavides Sepúlveda, A.M. and C.A. Blandón Jaramillo, *Modelo de sistema de gestión de seguridad de la información basado en la norma NTC ISO/IEC 27001 para instituciones públicas de educación básica de la comuna Universidad de la ciudad de Pereira Alejandra*. 2017.
- [17] Bornman, W.G., *Information security risk management: a holistic framework*, 2004, University of Johannesburg.
- [18] Refsdal, A., B. Solhaug, and K. Stolen, *Cyber-risk management*, in *Cyber-Risk Management* 2015, Springer. p. 33-47.
- [19] Zudin, R., *Analysis of information risk management methods*. University of Jyväskylä, 2014.
- [20] Carrillo Sánchez, J.P., *Guía y análisis de gestión de riesgos en la adquisición e implantación de equipamiento y servicios de tecnologías de información y comunicaciones para proyectos de alcance nacional*, 2012, Quito: EPN, 2012.
- [21] Pacheco Pozo, D.C., *Propuesta de un plan de contingencia de TI para la empresa LOGICIEL*, 2016, Quito, 2016.
- [22] Alcántara, M. and A. Melgar, *Risk management in information security: a systematic review*. Journal of Advances in Information Technology Vol, 2016. 7(1).
- [23] Shamala, P., et al., *Collective information structure model for information security risk assessment (ISRA)*. Journal of Systems and Information Technology, 2015. 17(2): p. 193-219.
- [24] Ford, M., *IT Risk Management Systems Play a Key Role in Sustaining and Promoting Business Growth*. University of Westminster, MSc Information Security, London, England, 2014.
- [25] Derakhshandeh, S. and N. Mikaeilvand, *New framework for comparing information security risk assessment methodologies*. Australian Journal of Basic and Applied Sciences, 2011. 5(9): p. 160-166.
- [26] Dehkoda, D., *Combining IRAM2 with Cost-Benefit Analysis for Risk Management: Creating a hybrid method with traditional and economic aspects*, 2018: Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology.
- [27] Duricu, A., *Data Protection Impact Assessment (DPIA) and Risk Assessment in the context of the General Data Protection Regulation (GDPR)*, E.a.S.E. Department of Computer Science, Luleå University of Technology Editor 2019.
- [28] Shamala, P. and R. Ahmad, *A proposed taxonomy of assets for information security risk assessment (ISRA)*. in *2014 4th World Congress on Information and Communication Technologies (WICT 2014)*. 2014. IEEE.
- [29] Wangen, G., *An initial insight into information security risk assessment practices*. in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*. 2016. IEEE.
- [30] Haythorn, M., *Information Security Risk Assessment Methods, Frameworks and Guidelines*. East Carolina University, 2013.
- [31] Pandey, S. and K. Mustafa, *Risk Assessment Framework (RAF)*. International Journal of Advanced Research in C. Sc., 2010. 1(3).
- [32] Rea-Guaman, A., et al. *Systematic Review: Cybersecurity Risk Taxonomy*. in *International Conference on Software Process Improvement*. 2017. Springer.
- [33] López, D., O. Pastor, and L.J.G. Villalba, *Concepto y Enfoques sobre el Análisis y la Gestión Dinámica del Riesgo en Sistemas de Información*. Actas de la XII Reunión Española de Criptología y Seguridad de la Información (RECSI 2012), Donostia-San Sebastián, España, 2012.
- [34] Ganin, A.A., et al., *Multicriteria decision framework for cybersecurity risk assessment and management*. Risk Analysis, 2017.
- [35] Smojver, S. *Selection of information security risk management method using analytic hierarchy process (ahp)*. in *Central European Conference on Information and Intelligent Systems*. 2011. Faculty of Organization and Informatics Varazdin.
- [36] Beckers, K., et al., *ISMS-CORAS: A structured method for establishing an ISO 27001 compliant information security management system*, in *Engineering Secure Future Internet Services and Systems* 2014, Springer. p. 315-344.
- [37] Shedden, P., et al. *Towards a knowledge perspective in information security risk assessments—an illustrative case study*. in *Proceedings of the 20th Australasian Conference on Information Systems*. 2009.
- [38] Rot, A. *Enterprise information technology security: risk management perspective*. in *Proceedings of the World Congress on Engineering and Computer Science*. 2009.
- [39] Saripalli, P. and B. Walters. *Quirc: A quantitative impact and risk assessment framework for cloud security*. in *2010 IEEE 3rd international conference on cloud computing*. 2010. Ieee.
- [40] Li, S., et al., *An improved information security risk assessments method for cyber-physical-social computing and networking*. IEEE Access, 2018. 6: p. 10311-10319.
- [41] Sicari, S., et al., *A risk assessment methodology for the Internet of Things*. Computer Communications, 2018. 129: p. 67-79.
- [42] Kitchenham, B., *Procedures for performing systematic reviews*. Keele, UK, Keele University, 2004. 33(2004): p. 1-26.
- [43] Brereton, P., et al., *Lessons from applying the systematic literature review process within the software engineering domain*. Journal of Systems and Software, 2007. 80(4): p. 571-583.
- [44] Budgen, D. and P. Brereton. *Performing systematic literature reviews in software engineering*. in *Proceedings of the 28th international conference on Software engineering*. 2006. ACM.
- [45] Biolchini, J., et al., *Systematic review in software engineering*. System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES, 2005. 679(05): p. 45.
- [46] Svatá, V. and M. Fleischmann, *IS/IT Risk Management in banking industry*. Acta oeconomica pragensia, 2011. 19(3): p. 42-60.
- [47] Mayer, N., P. Heymans, and R. Matulevicius. *Design of a Modelling Language for Information System Security Risk Managem*. 2007.
- [48] García, F.Y.H. and L.M.L. Moreta. *Maturity Model for the Risk Analysis of Information Assets based on Methodologies MAGERIT, OCTAVE y MEHARI; focused on Shipping Companies*. in *2018 7th International Conference On Software Process Improvement (CIMPS)*. 2018. IEEE.
- [49] Holguín García, F.Y., *Modelo de madurez para el análisis de riesgos de los activos de información basado en las metodologías MAGERIT, OCTAVE y MEHARI; con enfoque a empresas navieras*, 2018: Repositorio digital de la Universidad de Especialidades Espíritu Santo, UEES, Ecuador.
- [50] Gritzalis, D., et al., *Exiting the Risk Assessment maze: A meta-survey*. ACM Computing Surveys (CSUR), 2018. 51(1): p. 11.
- [51] Mrksic Kovacevic, S., *Smart homes from a Risk Management perspective*, 2018, University of Stavanger, Norway.
- [52] Wangen, G., C.V. Hallstensen, and E.A. Snekkenes, *A framework for estimating information security risk assessment method completeness: Core Unified Risk Framework*. 2017.
- [53] Wangen, G., C. Hallstensen, and E. Snekkenes, *A framework for estimating information security risk assessment method completeness*. International Journal of Information Security, 2018. 17(6): p. 681-699.
- [54] Novoa, H.A. and C.R. Barrera, *Metodologías para el análisis de riesgos en los sgsi*. Publicaciones e Investigación, 2015. 9: p. 73-86.
- [55] Santonja Lillo, J., *Análisis y correlación entre probabilidad e impacto de los riesgos*. Repositorio Institucional de la Universidad de Alicante, España., 2019.

- [56] Hashim, N.A., et al., *Risk Assessment Method for Insider Threats in Cyber Security: A Review*. Risk, 2018. **9**(11).
- [57] Bergvall, J. and L. Svensson, *Risk analysis review*, 2015: Linköpings Universitet, Linköping, Sweden.
- [58] Abbass, W., A. Baina, and M. Bellafkih. *Using EBIOS for risk management in critical information infrastructure*. in *2015 5th World Congress on Information and Communication Technologies (WICT)*. 2015. IEEE.
- [59] ENISA. (e). *Inventory of Risk Management / Risk Assessment Methods*. 2019 02/08/2019].
- [60] Pan, L. and A. Tomlinson, *A systematic review of information security risk assessment*. International Journal of Safety and Security Engineering, 2016. **6**(2): p. 270-281.
- [61] Shameli-Sendi, A., R. Aghababaei-Barzegar, and M. Cheriet, *Taxonomy of information security risk assessment (ISRA)*. Computers & Security, 2016. **57**: p. 14-30.
- [62] Ruan, K., *Introducing cybernomics: A unifying economic framework for measuring cyber risk*. Computers & Security, 2017. **65**: p. 77-89.
- [63] Madhavan, K. and R. ManickaChezian, *International Journal of Engineering Sciences & Research Technology a Study on Information Security and Risk Management in it Organizations*. International Journal OF Engineering Sciences & Research Technology (IJESRT). 2015.
- [64] Radanliev, P., et al., *Economic impact of IoT cyber risk-analysing past and present to predict the future developments in IoT risk analysis and IoT cyber insurance*. The Institution of Engineering and Technology (IET), England., 2018.
- [65] Acevedo, N. and C. Satizábal, *Risk management and prevention methodologies: a comparison*. Sistemas & Telemática, 2016. **14**(36): p. 39-58.
- [66] Satizábal, C., *Risk management and prevention methodologies: a comparison*. Sistemas & Telemática, vol.14, núm. 36, pp. 39-58, Universidad ICESI, Cali, Colombia, 2016.
- [67] Devia, G.A.V. and C.J. Pardo, *Hacia un modelo para la gestión de riesgos de TI en MiPyMEs: MOGRIT*. Sistemas & Telemática, 2014. **12**(30): p. 35-48.
- [68] Alhajri, R.M., et al. *Dynamic Interpretation Approaches for Information Security Risk Assessment*. in *2019 International Conference on Computer and Information Sciences (ICCIS)*. 2019. IEEE.
- [69] Korman, M., et al. *Overview of enterprise information needs in information security risk assessment*. in *2014 IEEE 18th International Enterprise Distributed Object Computing Conference*. 2014. IEEE.
- [70] Fulford, J.E., *What Factors Influence Companies' Successful Implementations of Technology Risk Management Systems?* Muma Business Review, 2017. **1**(13): p. 157-169.
- [71] Chen, F., *An Investigation and Evaluation of Risk Assessment Methods in Information systems*. Chalmers Univ. Technol. Goteborg, 2015: p. 1-83.
- [72] Pan, L., *Application of a Financial Quantitative Risk Model to Information Security Risk Assessment*. School of Mathematics and Information Security Royal Holloway, University London., 2018.
- [73] Nurse, J.R., S. Creese, and D. De Roure, *Security risk assessment in Internet of Things systems*. IT Professional, 2017. **19**(5): p. 20-26.
- [74] Agrawal, V., *A Comparative Study on Information Security Risk Analysis Methods*. Journal of Computers, 2017. **12**(1): p. 57-67.
- [75] Wangen, G.B., *Cyber Security Risk Assessment Practices: Core Unified Risk Framework*. Norges teknisk-naturvitenskapelige U., 2017.
- [76] MUKAMA, J., *Risk Analysis as a Security Metric for Industrial Control Systems*, in *Master's thesis in Computer Systems and Networks*. 2016, Chalmers University of Technology.
- [77] Oppliger, R., G. Pernul, and S. Katsikas, *New Frontiers: Assessing and Managing Security Risks*. Computer, 2017. **50**(4): p. 48-51.
- [78] Oppliger, R., *Quantitative risk analysis in information security management: a modern fairy tale*. IEEE Security & Privacy, 2015. **13**(6): p. 18-21.
- [79] Burnap, P., et al., *Determining and Sharing Risk Data in Distributed Interdependent Systems*. Computer, 2017. **50**(4): p. 72-79.
- [80] Rossebø, J.E., et al., *An enhanced risk-assessment methodology for smart grids*. Computer, 2017. **50**(4): p. 62-71.
- [81] Rossebø, J.E., F. Fransen, and E. Luijck. *Including threat actor capability and motivation in risk assessment for Smart GRIDs*. in *2016 Joint Workshop on Cyber-Physical Security and Resilience in Smart Grids (CPSR-SG)*. 2016. IEEE.

Luis Enrique Sánchez is PhD and MSc in Computer Science and is a Professor at the Universidad of Castilla-la Mancha (Ciudad Real, Spain), MSc in Information Systems Audit from the Polytechnic University of Madrid, and Certified Information System Auditor by ISACA. He is the Director of Professional Services and R&D departments of the company Sicaman Nuevas Tecnologías S.L. COIICLM board or committee member and responsible for the professional services committee. His research activities are management security system, security metrics, data mining, data cleaning, and business intelligence. He participates in the GSyA research group of the Department of Computer Science at the University of Castilla- LaMancha, in Ciudad Real (Spain).

Antonio Santos-Olmo is MSc in in Computer Science and is an Assistant Professor at the Escuela Superior de Informática de la Universidad de Castilla-La Mancha in Ciudad Real (Spain) (Computer Science Department, University of Castilla La Mancha, Ciudad Real, Spain), MSc in Information Systems Audit from the Polytechnic University of Madrid, and Certified Information System Auditor by ISACA. He is the Director of Software Factory departments of the company Sicaman Nuevas Tecnologías S.L. His research activities are management security system, security metrics, data mining, data cleaning, and business intelligence. He participates in the GSyA research group of the Department of Computer Science at the University of Castilla- LaMancha.

Victor Figueroa is Bachelor in Information Technology and MSc in Information Security. Is a Security Information Professor at the University of Siglo21 (Cordoba, Argentina). He is the Cybersecurity Director of Neuquen State (Argentina), working on development of Cybersecurity Information Policies, Risk Management and Incident Responses in Public Sector. He is currently a researcher in the field of Information Security Management Systems, and Information System Risk Analysis (ISRA).

David G. Rosado has an MSc and PhD. in Computer Science from the University of Málaga (Spain) and from the University of Castilla-La Mancha (Spain), respectively. His research activities are focused on security for Information Systems and Cloud Computing. He has published several papers in national and international conferences on these subjects, and he is co-editor of a book and chapter books. Author of several manuscripts in national and international journals (Information Software Technology, System Architecture, Network and Computer Applications, etc.). He is member of Program Committee of several conferences and workshops national and international such as ICEIS, ICCGI, CISIS, SBP, IAS, SDM, SECRIPT, COSE and international journals such as Internet Research, JNCA, KNOSYS, JKSU, and so on. He is a member of the GSyA research group of the Information Systems and Technologies Department at the University of Castilla-La Mancha, in Ciudad Real, Spain.

Eduardo Fernández-Medina holds a PhD. and an MSc. in Computer Science from the University of Sevilla. He is associate Professor at the Escuela Superior de Informática de la Universidad de Castilla-La Mancha at Ciudad Real (Spain), his research activity being in the field of security in databases, datawarehouses, web services and information systems, and also in security metrics. Fernández-Medina is co-editor of several books and chapter books on these subjects, and has several dozens of papers in national and international conferences (DEXA, CAISE, UML, ER, etc.). Author of several manuscripts in national and international journals (Information Software Technology, Computers And Security, Information Systems Security, etc.), he is director of the GSyA research group of the Information Systems and Technologies Department at the University of Castilla-La Mancha, in Ciudad Real, Spain.

Análisis de la Privacidad y de la Seguridad en Contratos Inteligentes

Antonio López Vivar, Ana Lucila Sandoval Orozco, and Luis Javier García Villalba, *Member, IEEE*

Resumen—Los contratos inteligentes han ganado mucha popularidad en los últimos tiempos ya que son una herramienta muy potente para el desarrollo de aplicaciones seguras descentralizadas y automáticas en multitud de campos sin necesidad de intermediarios o terceros de confianza. No obstante, debido a la naturaleza descentralizada de la cadena de bloques en la que se basan, se han puesto de manifiesto una serie de retos relacionados con la privacidad de la información que manejan dichos contratos, así como vulnerabilidades en su programación que, dadas sus particularidades, podrían tener (y ya han tenido) un impacto económico muy elevado. Este artículo proporciona una visión holística de los retos en materia de privacidad y seguridad asociados con los contratos inteligentes, así como del estado del arte de las herramientas y soluciones disponibles.

Palabras claves—Blockchain, privacy, security, smart contracts

I. INTRODUCCIÓN

Los contratos inteligentes está ganando popularidad en los últimos tiempos, aunque el concepto original es relativamente antiguo. La idea de los contratos inteligentes aparece en [1] y ha evolucionado hasta nuestros días, especialmente después de la introducción en 2009 de Bitcoin [2] y su cadena de bloques descentralizada. En pocas palabras, un contrato inteligente es un programa de ordenador que se ejecuta de forma descentralizada, modificando, si la ejecución es correcta, el estado global del sistema, que se almacena en una cadena de bloques. Aunque suele estar asociada a la plataforma Ethereum [3], hoy en día existen muchas plataformas que hacen uso de ellas, como por ejemplo Hyperledger o Corda entre muchas otras (en [4] puede consultarse una lista actualizada de todas las plataformas de contratos inteligentes existentes). En este artículo, nos centraremos en Ethereum, aunque mucho de los conceptos de seguridad son extrapolables al resto de plataformas.

En cuanto a las aplicaciones y casos de uso de los contratos inteligentes son muy variados pero su naturaleza descentralizada, velocidad, automatización, ausencia de intermediarios y transparencia los hacen particularmente adecuados en diferentes sectores, tales como: gestión de identidades [5] [6], voto electrónico [7] [8], servicios bancarios y financieros [9], gestión de herencias [10], cadena de suministro [11], IoT [12], juego en línea [3] e información médica [13] entre otros.

En el ejemplo del voto electrónico, sería posible mediante un contrato inteligente llevar la votación de unas elecciones y

el recuento de forma automatizada, transparente y sin requerir de una autoridad de confianza.

Este trabajo está estructurado en 5 secciones, siendo la primera la presente introducción. En la Sección II se introducen las de cadenas de bloques y los contratos inteligentes. La Sección III se centra en los problemas de privacidad de los contratos inteligentes y las herramientas. La sección IV se centra en las vulnerabilidades en la programación de los contratos. Finalmente, en la Sección V se presentan las conclusiones del presente trabajo.

II. CONTEXTO

Se presenta en esta sección una breve introducción de la cadena de bloques y los contratos inteligentes.

II-A. Tecnologías de Registro Distribuido

Una cadena de bloques es básicamente un registro replicado en múltiples nodos, llamados “mineros” que mantienen una copia (a veces no completa). Las operaciones, llamadas transacciones, se registran como en un libro de contabilidad, agrupándolas en bloques, que son añadidos a la cadena mediante un algoritmo de consenso basado habitualmente en una prueba de trabajo [14] Para realizar dicha prueba, cada uno de los nodos prepara un bloque candidato con un conjunto de transacciones nuevas pendientes de ser añadidas a la cadena de bloques al que le añade un contador y un puntero al último bloque de la cadena. Después cada nodo ejecuta sobre su bloque candidato un algoritmo de hash criptográfico (en el caso de Ethereum es KECCAK-256 [15]). Si el hash resultante es menor que un valor determinado, el bloque se considera minado y se propaga junto al hash calculado para ser verificado por el resto de nodos y que actualicen su copia de la cadena de bloques. En caso, contrario, si el hash resultante no es válido (lo más probable), se incrementa el contador del bloque y se vuelve a calcular el hash.

II-B. Contratos inteligentes

Como se ha mencionado, en la cadena de bloques se almacenan transacciones. En Ethereum, existen dos tipos de transacciones, las “normales”, donde un usuario A envía una cantidad de “Ether” (una moneda virtual) a la dirección de un usuario B. Las direcciones de usuario constan básicamente de un par de claves pública/privada y cada transacción que se genera va firmada digitalmente por el usuario que envía los fondos. Además de las direcciones de usuario, existen las direcciones de contrato, que como su nombre indican apuntan a un contrato inteligente. Para añadir un nuevo contrato inteligente a la cadena de bloques, un usuario tiene

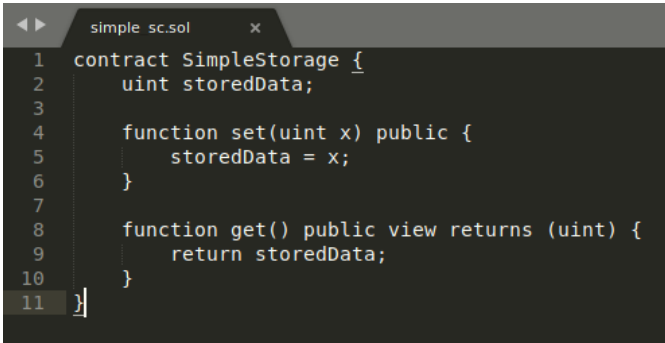
A. López Vivar, A. L. Sandoval Orozco and L. J. García Villalba. Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Ingeniería del Software e Inteligencia Artificial, Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, España. e-mail: {alopezvivar, asandoval, javiergv}@fdi.ucm.es.

que generar un tipo de transacción especial con una serie de datos, entre los que se incluye el código fuente del contrato ya compilado. Para ejecutar un contrato inteligente, un usuario tendrá que crear una transacción desde su dirección de usuario enviando Ether (y los parámetros que necesite el contrato) hacia la dirección del contrato. Si la ejecución del contrato es correcta, se modificará el estado global de Ethereum. En caso contrario, si la ejecución falla, se le cobrará al usuario el coste computacional usado del Ether que envió en la transacción, pero no habrá cambios en el estado global del sistema.

Los contratos inteligentes pueden escribirse en diferentes lenguajes de programación alto nivel, siendo el lenguaje más utilizado Solidity [16], con una sintaxis muy parecida a JavaScript. Este lenguaje es compilado a un bytecode que será ejecutado por la máquina virtual de Ethereum (EVM).

II-C. Un contrato inteligente de ejemplo

En la Figura 1 se presenta un ejemplo de un contrato inteligente muy básico que va a servir de ejemplo.



```

1 contract SimpleStorage {
2     uint storedData;
3
4     function set(uint x) public {
5         storedData = x;
6     }
7
8     function get() public view returns (uint) {
9         return storedData;
10    }
11 }

```

Figura 1. Contrato inteligente de ejemplo

Puede verse que se declara una variable de tipo entero sin signo que en Ethereum tiene 256 bits. Después se declara una función pública para asignar el valor de la variable anterior y otra función que devuelve el valor de la variable. Cada vez que algún usuario invoca este contrato y asigna un nuevo valor, éste se actualiza. Aunque en este ejemplo no se usan, hay mecanismos en los contratos inteligentes para restringir el acceso a determinadas funciones en función de quién ejecute el contrato inteligente.

Un dato importante a recordar es que antes de añadir un nuevo contrato a la cadena de bloques, los contratos se ejecutan de forma local en cada nodo minero y el resultado de la ejecución debe ser determinista para garantizar la coherencia de todo el sistema. Esta propiedad puede ser un problema a la hora de generar números aleatorios dentro de los contratos inteligentes. Por último, faltaría hablar del concepto de *gas*. Solidity es un lenguaje Turing-completo, por lo que existe la posibilidad de provocar bucles infinitos en la ejecución. Una manera de evitar esto, lo cual podría ser utilizado como forma de ataque de denegación de servicio contra la red de Ethereum, existe un mecanismo mediante el cual, cada una de las operaciones de bytecode posee un coste prefijado llamado *gas*. Cuando un usuario invoca un contrato, tiene que pasarle el precio unitario del *gas* y el límite de *gas* que está dispuesto a asumir. Si durante la ejecución del contrato, el *gas* alcanza el máximo que fijó el usuario, la ejecución se detiene lanzando

una excepción y el usuario que lanzó la ejecución pierde el *gas* multiplicado por el precio unitario. En caso de que la ejecución termine, después de pagado el precio del *gas* usado, el Ether sobrante será devuelto al usuario. Los mineros por su parte reciben unos honorarios para recompensarles por mantener la red. Sus honorarios están definidos por el *gas* y el precio del *gas*. Si el atacante intenta lanzar un ataque de denegación de servicio y elige un precio de *gas* acorde al mercado, los mineros ejecutarán el ataque pero el precio del ataque será muy alto. Por otro lado, si el atacante eligiera un precio del *gas* muy bajo, los mineros no incluirían su transacción en ningún bloque y por tanto no se ejecutaría el ataque.

III. PRIVACIDAD EN CONTRATOS INTELIGENTES

Los contratos inteligentes que se ejecutan en cadenas de bloques públicas, como es el caso de Ethereum, sufren de problemas de privacidad tanto en las transacciones, que quedan registradas en la cadena de bloques, como en el propio código fuente de los contratos. Si bien es cierto que las claves utilizadas para las transacciones de Ethereum no están asociadas a personas físicas concretas, hay varios estudios que analizando la cadena de bloques han podido correlacionar dichas transacciones y agruparlas [17], [18], [19]. Además, el propio código fuente de los contratos inteligentes, puede contener datos o claves de carácter privado y muy sensible, como es por ejemplo el caso de los contratos inteligentes asociados a dispositivos médicos. Dicho código fuente, por la propia arquitectura de la cadena de bloques se encuentra replicado para su ejecución en múltiples nodos repartidos por la red, por lo que se encuentran expuestos a fugas de información. En las secciones que vienen a continuación se hará una descripción de algunas de las herramientas más relevantes que existen actualmente para mitigar estos problemas de privacidad de los contratos inteligentes.

III-A. Hawk

Hawk es un framework para preservar la privacidad en los contratos inteligentes que funciona como una capa de abstracción para que el desarrollador escriba el código del contrato inteligente y después el compilador de Hawk creará tres programas a saber: el contrato inteligente que se ejecutará en la cadena de bloques, un programa que se ejecutará por lo usuarios y otro programa que será ejecutado por un manager (ver Figura 2).

Como puede verse en la Figura 2, un contrato de Hawk consta de dos partes diferenciadas: una parte privada donde se encuentran todos los datos sensibles que no queremos que sean visibles en la cadena de bloques y una parte pública donde se manejan datos de carácter menos sensible o públicos.

Para conseguir privacidad en la cadena de bloques, Hawk utiliza principalmente criptografía y pruebas de conocimiento cero para garantizar la integridad y el correcto funcionamiento de los contratos. Por otro lado, la privacidad a nivel de contrato no sólo se centra en usar criptografía para proporcionar confidencialidad si no que añade la figura del manager, una entidad que vigila comportamientos maliciosos de otros agentes, pero sin tener capacidad para modificar la correcta ejecución de los contratos inteligentes.

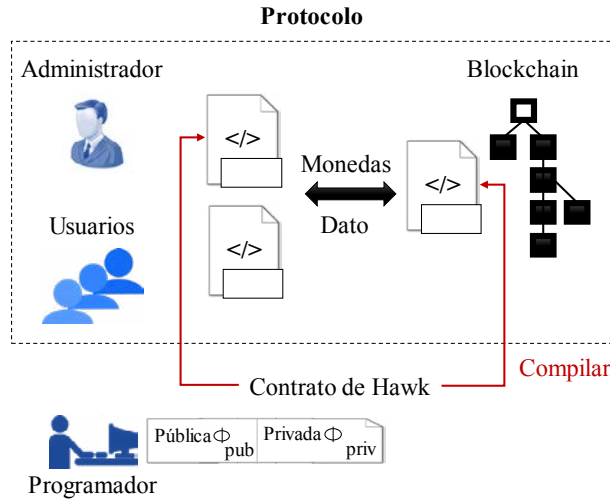


Figura 2. Esquema de funcionamiento de Hawk [20]

III-B. Ekiden

Ekiden [21], de forma parecida a Hawk, se trata de un entorno de ejecución segura de contratos inteligentes, pero a diferencia de Hawk que basa su área privada de ejecución en criptografía y pruebas de conocimiento cero, Ekiden apuesta por el uso de Entornos de Ejecución de Confianza [22] para ejecutar las partes más sensibles de los contratos inteligentes.

Como puede verse en la Figura 3, existen tres tipos de entidades en Ekiden: por un lado estarían los clientes, que serían los usuarios, los cuales pueden crear y ejecutar contratos. En segundo lugar se encontrarían los nodos de ejecución segura, que se encargarían de ejecutar los contratos y que deben disponer de una plataforma de ejecución en entorno de confianza. Por último, estarían los nodos de consenso, encargados de validar los resultados de ejecución de los nodos de computación y de añadir la información y cambios de estados del sistema a la cadena de bloques.

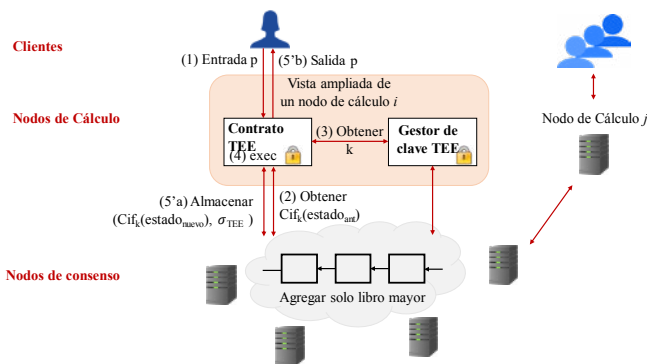


Figura 3. Arquitectura de Ekiden [21]

III-C. Enigma

Enigma [23] por su parte, utiliza computación multi-agente [24] para proveer seguridad en la ejecución de los contratos inteligentes. Los datos son segmentados y repartidos en varios nodos aleatoriamente entre varios nodos, de forma que ningún nodo tiene nunca la totalidad de la información. Enigma

separa la ejecución de ciertas partes intensivas en una cadena de bloques diferente de la principal, donde se guardan los resultados y las pruebas de verificación.

IV. SEGURIDAD EN CONTRATOS INTELIGENTES

Aunque la cadena de bloques está diseñada de forma segura y dicha seguridad está sustentada en algoritmos criptográficos ampliamente estudiados y probados, los contratos inteligentes como programas informáticos son susceptibles de contener vulnerabilidades de seguridad en su código, que dada su naturaleza inmutable y unido al hecho de que pueden funcionar en ámbitos financieros o de salud, supone una amenaza de seguridad grave, como ya ha sucedido en el pasado.

Un ejemplo famoso de ataque en Ethereum fue el *DAO attack* que se produjo en 2016 mediante el cual se robaron 3,6 millones de Ethers (el equivalente a unos 70 millones de dólares) aprovechando una vulnerabilidad en la programación de un contrato inteligente que controlaba el balance de una fundación. El contrato no tenía en cuenta la posibilidad de hacer llamadas recursivas que extraían dinero y no actualizaban el total, por lo que el hacker pudo realizar muchas de estas llamadas hasta vaciar los fondos. En [25] se hace un análisis detallado de este ataque.

En [26] se hace una clasificación de las vulnerabilidades existentes en los contratos inteligentes:

- Llamadas a lo desconocido: ocurre que algunas primitivas del lenguaje Solidity usadas para llamar a otras funciones o enviar Ether pueden sufrir de un efecto lateral de llamar a una función definida por defecto que tienen todos los contratos (y cuyo código podría ser desconocido para el llamador), en caso de no encontrarse la función llamada. Las primitivas afectadas por este efecto son:
 - CALL: es una primitiva utilizada para llamar a las funciones de un contrato inteligente (del mismo o de otro). Si la función que se le pasa como parámetro a la primitiva, no existe en el contrato, se ejecutará una función por defecto o *fallback*.
 - SEND: esta primitiva permite enviar Ether desde el contrato en ejecución a otro destinatario. Una vez que se ha enviado la cantidad de Ether, se ejecuta la función por defecto en el contrato de destino.
 - DELEGATECALL: esta primitiva es muy parecida a CALL a diferencia de que en ésta, se utiliza el contexto de variables del contrato llamador.
- Envío sin gas: al enviar Ether usando la primitiva SEND podría producirse una excepción por agotamiento de *gas* si el destinatario es un contrato tiene una función de *fallback* con mucho código.
- Desorden en las excepciones: en Solidity existen varias situaciones que pueden provocar que se dispare una excepción durante la ejecución, a saber: si la ejecución del contrato se queda sin *gas*, si se agota la pila de llamadas o en caso de que se lance la excepción de forma explícita llamando al comando *throw*. No obstante, Solidity no trata de la misma forma las excepciones si éstas se producen durante una llamada a una función de forma directa o usando la primitiva CALL. En el primer caso, la ejecución se detiene y cualquier efecto lateral se

revierte, incluyendo las transferencias de Ether. Pero si la excepción se produjo en el contexto de una llamada usando CALL, la excepción se propagará hacia arriba revirtiendo los efectos en los contratos llamados hasta alcanzar la llamada CALL, devolviendo *false* y continuará la ejecución a partir de ahí y consumiendo *gas*. Esta inconsistencia a la hora de manejar las excepciones puede dar lugar a vulnerabilidades.

- **Conversión de tipos:** aunque el compilador de Solidity puede detectar errores con los tipos, por ejemplo si una función espera un entero y es llamada pasándole una cadena, en el caso de definiciones de contratos o funciones con una determinada estructura, en caso de llamar a una función en un contrato, si el programador se equivoca y llama por error a otro contrato pero que contiene una función con la misma estructura esperada por el compilador se ejecutará la función y en caso de que no exista la función, se llamará a la función de *fallback*. En cualquier caso, no se lanzará ninguna excepción.
- **Reentrada:** esta es una vulnerabilidad muy conocida por su impacto. El programador puede pensar que una función no recursiva no puede ser re-llamada mientras se está ejecutando, pero esto no es siempre así, ya que podría darse el caso de que dentro de la función se llame a un contrato malicioso vacío que sólo contenga una función de *fallback* que vuelva a llamar a la función de la que viene. Por ejemplo, supongamos que tenemos un contrato así:

```
contract Bob {
    bool sent = false;
    function foo(address c) {
        if (!sent) {
            c.call.value(1)();
            sent = true;
        }
    }
}
```

La función *foo* recibe como parámetro la dirección de un contrato y si el testigo no está activado, envía 1 wei (la mínima unidad de Ether) al contrato *c*.

El problema viene si el contrato llamado es algo así:

```
contract Alice {
    function () {
        Bob(msg.sender).foo(this)
    }
}
```

En este caso, después de recibir el Ether, Alice llama su función de *fallback*, lo que a su vez vuelve a llamar a la función *foo* de Bob y como el testigo no está puesto a *true* volverá a transferir a Alice Ether y esto se repetirá hasta que se agote el *gas* o se alcance el límite de la pila de llamadas. Esta vulnerabilidad se utilizó el Ataque DAO comentado más arriba y en [25] se explica en detalle.

- **Secretos:** Solidity permite definir la visibilidad de los campos en los contratos como públicos o privados. Esto puede ser útil si se necesita ocultar determinada infor-

mación entre llamadas de contratos. Desgraciadamente este sistema no es efectivo ya que los cambios en campos privados tienen que enviarse a nodos mineros para que estos los metan en la cadena de bloques, la cual es pública. Aquellos contratos que necesiten ocultar información sensible tendrán que usar técnicas como las mencionadas en la Sección III

- **Estado impredecible:** todos los contratos inteligentes tienen un estado determinado por el valor de sus campos y su balance de Ether. Pero no se puede garantizar que el estado que tenía un contrato cuando hicimos una transacción hacia él será el mismo que cuando dicha transacción sea minada e incluida en la cadena de bloques. Es decir, podría ocurrir que antes de procesar nuestra transacción, otras transacciones hayan cambiado el estado del contrato de destino y además ser rápidos no nos garantiza nada porque los mineros pueden minar las transacciones en el orden que quieran. Se da otro problema añadido por la naturaleza de la cadena de bloques, y es que podría producirse un *fork* de la cadena si dos mineros consiguen minar un bloque válido al mismo tiempo. Esto provocaría que algunos mineros intentarían añadir su bloque en una de las dos cadenas y los otros en la otra. Llegado un momento la cadena más corta se desearía, perdiendo las transacciones contenidas en esta y cambiando el estado de los contratos a un estado indeterminado. Otro caso serían contratos que utilizan librerías dinámicas (un tipo especial de contratos que no pueden tener campos mutables). Ese tipo de contratos podrían cambiar de forma maliciosa para engañar a la víctima que los llamaría sin saber que han cambiado.
- **Números aleatorios:** la ejecución del código de la máquina virtual de Ethereum es determinista. Eso quiere decir, que el código ejecutado con las mismas entradas debe de producir las mismas salida en todos los nodos que lo ejecuten. Esto presenta un problema a la hora de generar número aleatorios. Para simular aleatoriedad, muchos contratos utilizan un generador de números aleatorios inicializado con la misma semilla para todos los mineros. Una opción muy utilizada por los programadores es utilizar como semilla el hash de un bloque determinado en el futuro. Al tratarse de un valor impredecible a priori, es una buena manera de inicializar el generador de números aleatorios. Sin embargo, como los mineros pueden elegir qué transacciones meter en los nuevos bloques, podrían conspirar para intentar conseguir alterar el funcionamiento del generador de números aleatorios.
- **Restricciones de tiempo:** muchas aplicaciones tienen restricciones de tiempo para operar. Habitualmente esas restricciones utilizan *timestamps*. En el caso de los contratos inteligentes, el programador puede obtener el *timestamp* de cuándo se minó el bloque, que es compartido por todas las transacciones del bloque. El problema es que los mineros en las primeras versiones del protocolo podían elegir el *timestamp* del bloque que iban a minar de forma arbitraria, lo que podía usarse para llevar a cabo ataques.
- **Bugs inmutables:** esto no es una vulnerabilidad en sí misma, sino la consecuencia de una propiedad de la ca-

dena de bloques. Todo el código fuente de los contratos inteligentes, incluidos aquellos que contienen *bugs* son inmutables una vez son minados y añadidos a la cadena de bloques, aunque si pueden ser bloqueados mediante la llamada a una función destructor.

- Pérdida de Ether: si el programador se equivoca al introducir la dirección para enviar Ether y esa dirección existe pero es una dirección huérfana que no pertenece a nadie ese Ether se perderá para siempre.
- Tamaño de pila: cada vez que un contrato llama a otro contrato la pila de llamada asociada aumenta en uno. El límite de la pila es 1024 y cuando se llega al límite se lanza una excepción. Hasta el 18 de octubre de 2016 era posible aprovecharse de esto para lanzar un ataque donde un usuario malicioso incrementaba el contador de la pila hasta casi agotarlo y entonces llamaba a la función de la víctima lo que lanzaba una excepción al agotar el límite de la pila. Si la víctima no tuvo en cuenta esto y no maneja correctamente la excepción, podría tener éxito el ataque. El impacto de esta vulnerabilidad hizo que se rediseñara Ethereum.

A fecha de creación de este documento existen multitud de herramientas para este propósito. A continuación se hará una pequeña descripción de los métodos o técnicas utilizados por las herramientas de análisis y reflejados en la tabla II y después de se hará una breve descripción de las herramientas.

IV-A. Métodos de análisis

- Ejecución simbólica: en vez de usar valores concretos para las variables se utilizan símbolos. Las operaciones sobre estos símbolos conducen a términos algebraicos, y las declaraciones condicionales dan lugar a fórmulas propositivas que caracterizan a las ramas. Una parte particular del código es alcanzable si la conjunción de fórmulas en el camino a esta parte es satisfactoria, lo que puede ser comprobado por los solucionadores de SMT.
- Interpretación asbtracta: mediante el uso de Árboles de Sintaxis Abstractos que se obtienen en una fase intermedia en el proceso de compilación de Solidity en bytecode

se analiza en el código en busca de vulnerabilidades.

- Reglas Horn: la lógica de Horn [27] está compuesta por una versión restringida de la lógica de primer orden donde todas las reglas son de la forma si-entonces que aunque limitada es computacionalmente universal y puede llevar a cabo los mismos cálculos que cualquier ordenador.
- Resolución de restricciones: mediante esta técnica las herramientas intentan resolver una serie de condiciones en el código para determinar si la ejecución podría tomar determinados caminos que conduzcan a potenciales vulnerabilidades.
- Verificación de modelo: este método se basa en una verificación automática de las propiedades de un sistema de estados finitos utilizando un modelo del sistema que se cruza con un conjunto de especificaciones.

IV-B. Herramientas

Oyente [28] es una de las primeras herramientas de seguridad que aparecieron. Ejecuta el bytecode EVM de los contratos inteligentes de forma simbólica y se centra en comprobar bloques de código susceptibles de ser vulnerables por su estructura, llamadas a contratos externos o una mala gestión de las excepciones de ejecución, por ejemplo la Figura 4. Disponible en Github [29] desde enero de 2016 (licencia GPL-3.0).

Remix-IDE es una extensión para el navegador web que permite escribir contratos inteligentes en el lenguaje Solidity. Muestra de forma gráfica distintos avisos de potenciales vulnerabilidades o fallos en el código a la vez que hace un análisis estático ligero. Disponible en Github [30] desde abril de 2016 (licencia MIT).

Solgraph [31] es una herramienta que permite visualizar el flujo ejecución de un contrato inteligente escrito en Solidity y ver de forma más fácil potenciales vulnerabilidades. Está disponible en Github [32] desde julio de 2016. *Porosity* [33] es una herramienta que desensambla el bytecode EVM y genera un diagrama de flujo. También permite descompilar el bytecode y convertirlo a código fuente. Disponible en Github [34] desde febrero de 2017.

Tabla I
VULNERABILIDADES DETECTADAS POR LAS HERRAMIENTAS

		Oyente	Remix-IDE	Solgraph	Porosity	Manticore	SmartCheck	FSolidM	Mythril	ContractLarva	E-EVM	SolMet	Vandal	EthIR	MAIAN	Erays	Rattle	Osiris	Securify	Ethertrust
Solidity	Llamadas a lo desconocido	x	✓	x	x	✓	✓	x	✓	x	x	x	x	x	✓	x	x	x	✓	x
	Envío sin gas	x	✓	x	x	✓	✓	x	✓	x	x	x	x	x	✓	x	x	x	✓	x
	Desorden en excepciones	x	✓	x	x	✓	✓	x	✓	x	x	x	x	x	✓	x	x	x	✓	x
	Conversión tipos	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	Reentrada	✓	✓	x	✓	✓	✓	x	✓	x	x	x	✓	x	x	x	x	x	✓	x
	Conversión tipos	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	Revelación de secretos	x	✓	✓	x	x	✓	x	x	x	x	x	x	x	x	x	x	x	x	x
Blockchain	Estado impredecible	x	x	x	x	x	✓	x	x	x	x	x	x	x	x	x	x	x	✓	x
	Números aleatorios	x	x	x	x	x	x	x	✓	x	x	x	x	x	x	x	x	x	x	x
	Restricciones de tiempo	✓	✓	x	x	✓	✓	x	✓	x	x	x	x	x	x	x	x	x	x	x
EVM	Bugs inmutables	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	Pérdida de Ether	x	x	x	x	x	✓	x	x	x	x	x	x	x	✓	x	x	x	✓	x
	Tamaño de pila	✓	x	x	x	x	x	x	✓	x	x	x	x	x	x	x	x	x	x	x

Manticore [35] es una herramienta de ejecución simbólica que en el caso de la EVM, analiza la ejecución del código del contrato inteligente en busca de las vulnerabilidades mas típicas apoyándose del solucionador SMT Z3 [36]. La herramienta está disponible en Github [37] desde febrero de 2017 (licencia AGPL-3.0).

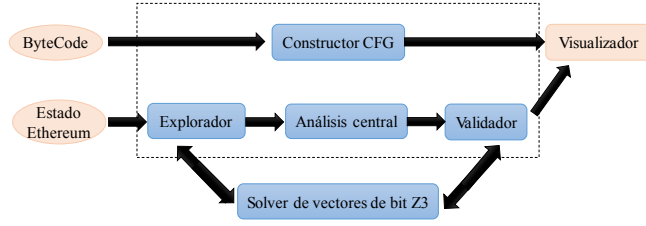


Figura 4. Esquema de funcionamiento de Oyente [28]

SmartCheck [38] busca patrones de vulnerabilidades conocidas en el código fuente. Para ello utiliza un archivo XML con el código fuente en forma de árbol y explora caminos que pueden llevar a vulnerabilidades durante la ejecución. Disponible en Github [39] desde mayo de 2017 (licencia GPL-3.0).

FsolidM [40] utiliza una máquina de estados finita para permitir al usuario definir el comportamiento del contrato inteligente y a partir de ésta generar de forma automática el código en Solidity del contrato inteligente. Está disponible en Github [41] desde septiembre de 2017 (licencia MIT).

Mythril [42], es una herramienta muy parecida a Manticore, que permita la ejecución simbólica del bytecode EVM y que genera un grafo de control de flujo. Todo esto permite detectar una serie de tipos de vulnerabilidades. Desarrollado por la empresa ConsenSys, se encuentra disponible en Github [43] desde septiembre de 2017.

contractLarva [44] Es una herramienta que verifica los contratos inteligentes en tiempo de ejecución. Para ello, al compilar el contrato, añade al código fuente original, una serie de instrucciones encargadas de velar durante la ejecución de que no se producen vulnerabilidades y/o evitarlas dentro de lo posible sin alterar el flujo de ejecución normal del

contrato. Está disponible en Github [45] desde diciembre de 2017 (licencia Apache-2.0).

E-EVM [46] es una herramienta que permite la ejecución del bytecode EVM de forma visual, creando un grafo de flujo y mostrando información sobre la pila de ejecución, a modo de depurador. Está disponible en Github [47] desde enero de 2018.

SolMet [48] se trata de una herramienta para calcular la complejidad del código fuente de un contrato escrito en Solidity que utiliza un parser para generar un árbol de sintaxis abstracta donde evalúa una serie de métricas para medir la complejidad. Disponible en Github [49] desde febrero de 2018.

Vandal [50] es un desensamblador y descompilador de Solidity que genera un grafo de control de flujo que puede visualizarse como una página web HTML. Además permite especificar los análisis de seguridad de forma lógica usando el lenguaje Soufflé [51]. Disponible en Github [52] desde febrero de 2018.

EthIR [44] es un framework que transforma el código fuente de los contratos inteligentes a un lenguaje intermedio. Está basado en la herramienta Oyente pero utiliza una representación basada en reglas. Se encuentra disponible en Github [53] desde marzo de 2018 (licencia GPL-3.0).

MAIAN [54] es una herramienta muy parecida a Oyente, pero que la amplía teniendo en cuenta los ataques que requieren de varias transacciones. Se apoya en el uso de una cadena de bloques privada para testeo y así mitigar los falsos positivos. Al igual que otras muchas herramientas utiliza el solucionador SMT Z3 para buscar caminos en la ejecución que lleven a potenciales vulnerabilidades. Está disponible en Github [55] desde marzo de 2018 (licencia MIT).

Erays [56] es un desensamblador de bytecode EVM que genera un archivo PDF con pseudocódigo de las rutinas presentes en los contratos inteligentes. Se encuentra disponible en Github [57] desde agosto de 2018 (licencia MIT).

Rattle [58] es un framework que hace análisis estático binario del bytecode EVM realizando un desensamblado que elimina instrucciones no necesarias para entender el funcionamiento del código fuente del contrato, aunque no detecta vulnerabilidades por sí mismo. Disponible en Github [59] desde agosto de 2018 (licencia GPL-3.0).

Tabla II
RESUMEN DE LAS HERRAMIENTAS DE ANÁLISIS

		Oyente	Remix-IDE	Solgraph	Porosity	Manticore	SmartCheck	FSolidM	Mythril	ContractLarva	E-EVM	SolMet	Vandal	EthIR	MAIAN	Erays	Rattle	Osiris	Securify	Ethertrust
Nivel	Bytecode	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Solidity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tipo	Análisis dinámico	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Análisis estático	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Algoritmo	Ejecución simbólica	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Interpretación abstracta	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Instrumentación de código	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Reglas Horn	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Resolución de restricciones	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Verificación de modelo	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Osiris [60] es una herramienta especializada en detectar vulnerabilidades relacionadas con números enteros dentro de los contratos inteligentes de Ethereum. Funciona ampliando la funcionalidad de Oyente analizando el flujo de ejecución para distinguir desbordamientos considerados benignos de aquellos malignos. Disponible en Github [61] desde septiembre de 2018.

Securify [62] realiza primero un desensamblado del bytecode EVM. Después, el código desensamblado lo descompila en un lenguaje intermedio con el que crea un conjunto de reglas DataLog como las mostradas en la figura 6 que usará para buscar la violación de patrones y así encontrar vulnerabilidades. Disponible en Github [63] desde septiembre de 2018 (licencia Apache-2.0).

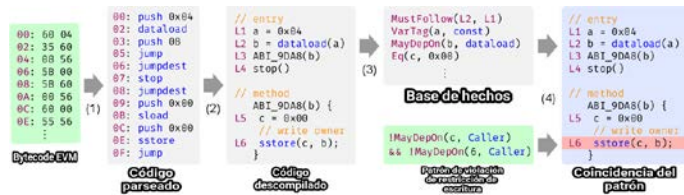


Figura 5. Análisis de código Securify [62]

EtherTrust [64] es un framework que basa su funcionamiento en traducir el bytecode EVM los contratos inteligentes en cláusulas de Horn que junto al solucionador SMT Z3 verifica que el código del contrato no presente potenciales vulnerabilidades aunque no las detecta. Está disponible en Github [65] desde agosto de 2019 (licencia GPL-3.0).

En la tabla I basada en la taxonomía de vulnerabilidades de [26] puede verse un resumen de todas las vulnerabilidades que las herramientas de seguridad analizadas detectan. Algunas de ellas, como por ejemplo FSolidM o ContractLarva no detectan ninguna porque no es su función.

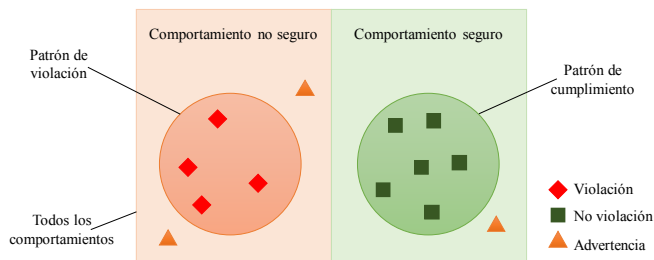


Figura 6. Sistema de patrones basados en reglas que utiliza Securify [62]

V. CONCLUSIONES Y TRABAJOS FUTUROS

La seguridad y privacidad de contratos inteligentes es un campo de investigación relativamente nuevo, pero con este documento pretendemos dejar patente que ya existen multitud de herramientas y soluciones, algunas de ellas bastante maduras. Es cierto, que con tanta variedad de soluciones, el desarrollador de contratos inteligentes puede a priori sentirse confundido y no saber por dónde empezar, aunque trabajos como este podrían ser un buen punto de partida.

En el caso de las herramientas de análisis de vulnerabilidades, aunque pueda parecer en principio que existe mucha

variedad, muchas de ellas comparten algoritmos y técnicas de análisis y otras son evoluciones de herramientas más antiguas. Además, a la hora de desarrollar nuevas herramientas, existen ya librerías que aceleran el proceso como el desensamblador de código de Ethereum de la Fundación Ethereum.

Por último, como trabajo futuro, quedaría pendiente el estudio de todas aquellas herramientas no disponibles públicamente (y/o cuyo código fuente no es público) y que no se han incluido en este trabajo.

AGRADECIMIENTOS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700326. Website: <http://ramses2020.eu>. This paper has also received funding from THEIA (Techniques for Integrity and authentication of multimedia files of mobile devices) UCM project (FEI-EU-19-04).



REFERENCIAS

- [1] N. Szabo, "Formalizing and securing relationships on public networks," *First Monday*, vol. 2, no. 9, 1997.
- [2] S. Nakamoto *et al.*, "Bitcoin: A peer-to-peer electronic cash system," 2008.
- [3] V. Buterin *et al.*, "A next-generation smart contract and decentralized application platform," *white paper*, vol. 3, p. 37, 2014.
- [4] Coinlore, "Smart contract platforms."
- [5] M. Al-Bassam, "Scpki: a smart contract-based pki and identity system," in *Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts*. ACM, 2017, pp. 35–40.
- [6] P. Dunphy and F. A. Petitcolas, "A first look at identity management schemes on the blockchain," *IEEE Security & Privacy*, vol. 16, no. 4, pp. 20–29, 2018.
- [7] P. McCorry, S. F. Shahandashti, and F. Hao, "A smart contract for boardroom voting with maximum voter privacy," in *International Conference on Financial Cryptography and Data Security*. Springer, 2017, pp. 357–375.
- [8] N. Kshetri and J. Voas, "Blockchain-enabled e-voting," *IEEE Software*, vol. 35, no. 4, pp. 95–99, 2018.
- [9] G. W. Peters and E. Panayi, "Understanding modern banking ledgers through blockchain technologies: Future of transaction processing and smart contracts on the internet of money," in *Banking beyond banks and money*. Springer, 2016, pp. 239–278.
- [10] P. Sreehari, M. Nandakishore, G. Krishna, J. Jacob, and V. Shibu, "Smart will converting the legal testament into a smart contract," in *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*. IEEE, 2017, pp. 203–207.
- [11] T. Bocek, B. B. Rodrigues, T. Strasser, and B. Stiller, "Blockchains everywhere-a use-case of blockchains in the pharma supply-chain," in *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE, 2017, pp. 772–777.
- [12] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the internet of things," *Ieee Access*, vol. 4, pp. 2292–2303, 2016.
- [13] N. Rifi, E. Rachkidi, N. Agoulmine, and N. C. Taher, "Towards using blockchain technology for ehealth data access management," in *2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME)*. IEEE, 2017, pp. 1–4.
- [14] A. Gervais, G. O. Karame, K. Wüst, V. Glykantzis, H. Ritzdorf, and S. Capkun, "On the security and performance of proof of work blockchains," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. ACM, 2016, pp. 3–16.
- [15] G. Bertoni, J. Daemen, M. Peeters, and G. Van Assche, "Keccak specifications," *Submission to nist (round 2)*, pp. 320–337, 2009.
- [16] C. Dannen, *Introducing Ethereum and Solidity*. Springer, 2017.
- [17] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," *Future Generation Computer Systems*, 2017.
- [18] F. Reid and M. Harrigan, "An analysis of anonymity in the bitcoin system," in *Security and privacy in social networks*. Springer, 2013, pp. 197–223.

- [19] M. Möser, K. Soska, E. Heilman, K. Lee, H. Heffan, S. Srivastava, K. Hogan, J. Hennessey, A. Miller, A. Narayanan *et al.*, “An empirical analysis of traceability in the monero blockchain,” *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 3, pp. 143–163, 2018.
- [20] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, “Hawk: The blockchain model of cryptography and privacy-preserving smart contracts,” in *2016 IEEE symposium on security and privacy (SP)*. IEEE, 2016, pp. 839–858.
- [21] R. Cheng, F. Zhang, J. Kos, W. He, N. Hynes, N. Johnson, A. Juels, A. Miller, and D. Song, “Ekiden: A platform for confidentiality-preserving, trustworthy, and performant smart contracts,” in *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2019, pp. 185–200.
- [22] G. Arfaoui, S. Gharout, and J. Traoré, “Trusted execution environments: A look under the hood,” in *2014 2nd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*. IEEE, 2014, pp. 259–266.
- [23] G. Zyskind, O. Nathan, and A. Pentland, “Enigma: Decentralized computation platform with guaranteed privacy,” *arXiv preprint arXiv:1506.03471*, 2015.
- [24] Y. Lindell, “Secure multiparty computation for privacy preserving data mining,” in *Encyclopedia of Data Warehousing and Mining*. IGI Global, 2005, pp. 1005–1009.
- [25] M. I. Mehar, C. L. Shier, A. Giambattista, E. Gong, G. Fletcher, R. Sanyahie, H. M. Kim, and M. Laskowski, “Understanding a revolutionary and flawed grand experiment in blockchain: the dao attack,” *Journal of Cases on Information Technology (JCIT)*, vol. 21, no. 1, pp. 19–32, 2019.
- [26] N. Atzei, M. Bartoletti, and T. Cimoli, “A survey of attacks on ethereum smart contracts (sok),” in *International Conference on Principles of Security and Trust*. Springer, 2017, pp. 164–186.
- [27] A. Gurfinkel, T. Kahsai, A. Komuravelli, and J. A. Navas, “The seahorn verification framework,” in *International Conference on Computer Aided Verification*. Springer, 2015, pp. 343–361.
- [28] L. Luu, D.-H. Chu, H. Olickel, P. Saxena, and A. Hobor, “Making smart contracts smarter,” in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. ACM, 2016, pp. 254–269.
- [29] Melon Project, “Oyente,” <https://github.com/melonproject/oyente>, 2016.
- [30] Ethereum Foundation, “Remix-ide,” <https://github.com/ethereum/remix-ide>, 2018.
- [31] R. Revere, “Solgraph,” <https://github.com/raineorshine/solgraph>, 2018.
- [32] R. Revere, “Solgraph,” <https://github.com/raineorshine/solgraph>, 2016.
- [33] M. Suiche, “Porosity: A decompiler for blockchain-based smart contracts bytecode,” *DEF con*, vol. 25, p. 11, 2017.
- [34] Comae Technologies, “Porosity,” <https://github.com/comaeio/porosity>, 2017.
- [35] M. Mossberg, F. Manzano, E. Hennenfent, A. Groce, G. Grieco, J. Feist, T. Brunson, and A. Dinaburg, “Manticore: A user-friendly symbolic execution framework for binaries and smart contracts,” *arXiv preprint arXiv:1907.03890*, 2019.
- [36] L. De Moura and N. Björner, “Z3: An efficient smt solver,” in *International conference on Tools and Algorithms for the Construction and Analysis of Systems*. Springer, 2008, pp. 337–340.
- [37] T. of Bits, “Manticore,” <https://github.com/trailofbits/manticore>.
- [38] S. Tikhomirov, E. Voskresenskaya, I. Ivanitskiy, R. Takhaviev, E. Marchenko, and Y. Alexandrov, “Smartcheck: Static analysis of ethereum smart contracts,” in *2018 IEEE/ACM 1st International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB)*. IEEE, 2018, pp. 9–16.
- [39] SmartDec, “Smartcheck,” <https://github.com/smartdec/smartcheck>, 2017.
- [40] A. Mavridou and A. Laszka, “Tool demonstration: Fsolidm for designing secure ethereum smart contracts,” in *International Conference on Principles of Security and Trust*. Springer, 2018, pp. 270–277.
- [41] A. Mavridou and A. Laszka, “Fsolidm,” <https://github.com/anmavrid/smart-contracts>, 2017.
- [42] B. Mueller, “Smashing smart contracts,” in *9th HITB Security Conference*, 2018.
- [43] ConsenSys, “Mythril,” <https://github.com/ConsenSys/mythril-classic>, 2017.
- [44] E. Albert, P. Gordillo, B. Livshits, A. Rubio, and I. Sergey, “Ethir: A framework for high-level analysis of ethereum bytecode,” in *International Symposium on Automated Technology for Verification and Analysis*. Springer, 2018, pp. 513–520.
- [45] G. Pace, “contractlarva,” <https://github.com/gordonpace/contractLarva>, 2017.
- [46] R. Norvill, B. B. F. Pontiveros, R. State, and A. Cullen, “Visual emulation for ethereum’s virtual machine,” in *NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2018, pp. 1–4.
- [47] pisocrob, “E-vm,” <https://github.com/pisocrob/E-EVM>, 2018.
- [48] P. Hegedus, “Towards analyzing the complexity landscape of solidity based ethereum smart contracts,” *Technologies*, vol. 7, no. 1, p. 6, 2019.
- [49] P. Hegedus, “Solmet,” <https://github.com/chicxurug/SolMet-Solidity-parser>, 2018.
- [50] L. Brent, A. Jurisevic, M. Kong, E. Liu, F. Gauthier, V. Gramoli, R. Holz, and B. Scholz, “Vandal: A scalable security analysis framework for smart contracts,” *arXiv preprint arXiv:1809.03981*, 2018.
- [51] H. Jordan, B. Scholz, and P. Subotić, “Soufflé: On synthesis of program analyzers,” in *International Conference on Computer Aided Verification*. Springer, 2016, pp. 422–430.
- [52] Smart Contract Research (USYD), “Vandal,” <https://github.com/usyd-blockchain/vandal>, 2018.
- [53] P. Gordillo, “Ethir,” <https://github.com/costa-group/EthIR>, 2018.
- [54] I. Nikolić, A. Kolluri, I. Sergey, P. Saxena, and A. Hobor, “Finding the greedy, prodigal, and suicidal contracts at scale,” in *Proceedings of the 34th Annual Computer Security Applications Conference*. ACM, 2018, pp. 653–663.
- [55] MAIAN-tool, “Maian,” <https://github.com/MAIAN-tool/MAIAN>, 2018.
- [56] H. Gans, “Poole and plans: Erays on urban solutions and problems,” 1968.
- [57] teamnrg, <https://github.com/teamnrg/erays>, 2018.
- [58] R. Stortz, “Rattle - an Ethereum EVM binary analysis framework,” <https://github.com/crytic/rattle>, 2018.
- [59] Trail of Bits, “Rattle,” <https://github.com/trailofbits/rattle>, 2018.
- [60] C. F. Torres, J. Schütte *et al.*, “Osiris: Hunting for integer bugs in ethereum smart contracts,” in *Proceedings of the 34th Annual Computer Security Applications Conference*. ACM, 2018, pp. 664–676.
- [61] C. Ferreira, “Osiris,” <https://github.com/christoftorres/Osiris>, 2018.
- [62] P. Tsankov, A. Dan, D. Drachsler-Cohen, A. Gervais, F. Buenzli, and M. Vechev, “Securify: Practical security analysis of smart contracts,” in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2018, pp. 67–82.
- [63] SRI Lab, “Securify,” <https://github.com/eth-sri/securify>, 2018.
- [64] I. Grishchenko, M. Maffei, and C. Schneidewind, “Ethertrust: Sound static analysis of ethereum bytecode,” *Technische Universität Wien, Tech. Rep.*, 2018.
- [65] SecPriv, “Ethertrust,” <https://github.com/SecPriv/EtherTrust>, 2019.

Antonio López Vivar received his Computer Engineering degree at Universidad Carlos III of Madrid (2011) and Master’s Degree in Security of Information and Communication Technologies at Universidad Europea of Madrid (2015). Currently he is a Ph.D. student in the Department of Software Engineering and Artificial Intelligence of the Faculty of Computer Science and Engineering at the Universidad Complutense de Madrid (UCM) and Member of the Complutense Research Group GASS (Group of Analysis, Security and Systems, <http://gass.ucm.es>). working as Research Support Staff in the Group of Analysis, Security and Systems (GASS) of Universidad Complutense de Madrid. His research interests are: blockchain, cryptocurrencies, computer forensics and cybersecurity.

Ana Lucila Sandoval Orozco was born in Chivolo, Magdalena, Colombia in 1976. She received a Computer Science Engineering degree from the Universidad Autónoma del Caribe (Colombia) in 2001. She holds a Specialization Course in Computer Networks (2006) from the Universidad del Norte (Colombia), and holds a M.Sc. in Research in Computer Science (2009) and a Ph.D. in Computer Science (2014), both from the Universidad Complutense de Madrid (Spain). She is currently a postdoctoral researcher and member of the Research Group GASS (Group of Analysis, Security and Systems, <http://gass.ucm.es>) at Universidad Complutense de Madrid (Spain). Her main research interests are coding theory, information security and its applications.

Luis Javier García Villalba received a Telecommunication Engineering degree from the Universidad de Málaga (Spain) in 1993 and holds a Ph.D. in Computer Science (1999) from the Universidad Politécnica de Madrid (Spain). Visiting Scholar at COSIC (Computer Security and Industrial Cryptography, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, Belgium) in 2000 and Visiting Scientist at IBM Research Division (IBM Almaden Research Center, San Jose, CA, USA) in 2001 and 2002, he is currently Associate Professor of the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM) and Head of Complutense Research Group GASS (Group of Analysis, Security and Systems) which is located in the Faculty of Computer Science and Engineering at the UCM Campus.

His professional experience includes the management of both national and international research projects and both public (Spanish Ministry of R&D, Spanish Ministry of Defence, Horizon 2020 - European Commission, . . .) and private financing (Hitachi, IBM, Nokia, Safelayer Secure Communications, TB Solutions Security, . . .). Author or co-author of numerous international publications is editor or guest editor of numerous journals such as Entropy MPDI, Future Generation Computer Systems (FGCS), Future Internet MDPI, IEEE Latin America Transactions, IET Communications (IET-COM), IET Networks (IET-NET), IET Wireless Sensor Systems (IET-WSS), International Journal of Ad Hoc and Ubiquitous Computing (IJAHUC), International Journal of Multimedia and Ubiquitous Engineering (IJMUE), Journal of Supercomputing, Sensors MDPI, etc.

Revisión Sistemática de Análisis de Riesgos Asociativos y Jerárquicos. Periodo 2014 – 2019.

A. Santos-Olmo, L. E. Sánchez, E. Álvarez, D.G. Rosado, E. Fernandez-Medina

Resumen – La sociedad de la información cada vez depende más de los Sistemas de Gestión y Análisis del Riesgo al que se encuentran sometidos sus principales activos de información, y poder disponer de estos sistemas ha llegado a ser vital para la evolución de las PYMES. Sin embargo, este tipo de compañías requiere que estos sistemas estén adaptados a sus especiales características, y teniendo en cuenta la existencia de riesgos derivados no sólo de la propia PYME, sino riesgos externos de otras empresas que colaboran con ella, mediante relaciones de asociatividad y jerarquía. De esta forma, obtendremos un análisis de riesgo de mayor calidad (y reduciendo su coste) empleando conceptos avanzados como “Algoritmos asociativos” y “Redes sociales empresariales”. En este artículo presentamos los resultados obtenidos tras aplicar el método de investigación “Revisión Sistemática de la Literatura” de las propuestas científicas orientadas a análisis de riesgos TIC Asociativos y Jerárquicos, publicadas en los últimos 5 años.

Palabras clave — Cibersecurity, Information Systems Security Management, ISMS, Risk Analysis, SME, ISO27001, ISO27002, ISO27005, Magerit.

I. INTRODUCCIÓN

Para las empresas, es muy importante implantar controles de seguridad que les permitan conocer y controlar los riesgos a los que pueden estar sometidas [1-3]. Pero la implantación de estos controles no es suficiente, siendo necesarios sistemas que gestionen la seguridad a lo largo del tiempo, de modo que les permitan reaccionar ágilmente ante nuevos riesgos, vulnerabilidades, amenazas, etc. [4]. Sin embargo, la mayor parte de las empresas tienen sistemas de seguridad caóticos creados sin unas guías adecuadas, sin documentación y con recursos insuficientes [5]. Los controles clásicos se muestran por sí solos insuficientes para dar unas mínimas garantías de seguridad, en especial en sectores como el de la salud [6, 7]. Las herramientas de seguridad existentes en el mercado ayudan a solucionar parte de los problemas de seguridad, pero nunca afrontan el problema de una manera global e integrada. Por último, la enorme diversidad de estas herramientas y su falta de integración suponen un enorme coste en recursos para poderlas gestionar.

Por lo tanto, a pesar de que la realidad ha demostrado que para que las empresas puedan utilizar las tecnologías de la

información y las comunicaciones con garantías es necesario disponer de guías, métricas y herramientas que les permitan conocer en cada momento su nivel de seguridad y las vulnerabilidades que aún no han sido cubiertas [8], el nivel de implantación con éxito de estos sistemas realmente es muy bajo [9]. Este problema se acentúa especialmente en el caso de las pequeñas y medianas empresas, que cuentan con la limitación adicional de no tener recursos humanos y económicos suficientes para realizar una adecuada gestión [5].

Algunos autores [10, 11] sugieren la realización de un análisis de riesgos como parte fundamental en la PYME, ya que deben tener en cuenta que el valor y la sanción de los datos robados o filtrados en una pequeña organización es el mismo que para una grande, y por tanto debe tener controlado el valor y los riesgos a los que esos activos están sometidos. Otros autores [12] proponen la necesidad de desarrollar un nuevo modelo de análisis de riesgos orientándolo directamente a las PYMES.

Algunos autores [13] sugieren que no es suficiente con aplicar un enfoque basado en análisis y gestión de riesgos sino que, además de identificar y eliminar riesgos, también esta actividad se ha de realizar de manera eficiente, ahorrando dinero, consecuencia directa de una correcta gestión de la seguridad [14, 15].

Además, en una época en la que la colaboración es vital en la situación actual del mercado, es necesario contemplar también el riesgo derivado de la relación de la empresa con su entorno, sus circunstancias (variantes en cada momento) y con otras empresas, bien partners tecnológicos, bien como terceras partes en algún servicio que realice la empresa o bien como co-participantes en proyectos multi-empresa.

El tratamiento de estos riesgos de tipo asociativo adquiere también especial relevancia con la aparición del Cloud Computing, que ha alterado drásticamente la percepción de las arquitecturas de infraestructura de Sistemas de Información, con el consiguiente deterioro de gran parte de la eficacia de los mecanismos tradicionales de protección [16].

Añadido a este tipo de riesgo, también es necesario gestionar los riesgos de carácter vertical en la jerarquía de empresa, donde la actividad de una empresa filial puede afectar a la empresa matriz, y viceversa.

De esta manera, el objetivo principal de este trabajo es realizar una revisión sistemática de los modelos y metodologías existentes o en desarrollo para el análisis y gestión de riesgos, contemplando riesgos de carácter asociativo y jerárquico, y con orientación a PYMES.

En este artículo, vamos a llevar a cabo una revisión sistemática (RS) de la literatura existente en relación con las investigaciones en el campo de los Análisis de Riesgos, no

A. Santos-Olmo, Departamento I+D+i, Sicaman Nuevas Tecnologías y Marisma Shield, Tomelloso (Ciudad Real), España, Asolmo@sicaman-nt.com
L. E. Sánchez, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, Luisenrique@sanchezcespso.org
E. Álvarez, Fundación In-Nova, Toledo, España, Ealvarez@in-nova.org
D.G. Rosado, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, David.Grosado@uclm.es
E. Fernandez-Medina, Grupo de Investigación GSyA, Universidad de Castilla-la Mancha, Ciudad Real, España, Eduardo.FdezMedina@uclm.es

sólo con el fin de resumir las pruebas existentes en relación con este tema, sino también para proporcionar un marco en el que posicionar adecuadamente nuevas líneas de investigación.

Esta revisión sistemática se va a llevar a cabo mediante el uso de las directrices para las revisiones sistemáticas propuestas por Kitchenham [17-19], lo que se ha mostrado apropiado para investigadores de ingeniería del software. Usaremos también un modelo de protocolo de revisión desarrollado por Biolchini et al. [20], que facilita la planificación y ejecución de las revisiones sistemáticas en la ingeniería del software.

El resto del artículo se estructura de la siguiente forma: En la Sección 2 definiremos la pregunta de investigación. En la Sección 3 se explica el método de revisión, que se basa en el protocolo de investigación, y es aquí donde se definirá la estrategia de búsqueda y la selección de estudios. En la Sección 4 definiremos los datos a extraer y que se presentarán en el resumen de los estudios relevantes. En Sección 5 presentaremos los resultados de la revisión y un análisis de los mismos. Finalmente, en la última sección describiremos las principales conclusiones obtenidas.

II. PLANIFICACIÓN DE LA REVISIÓN

En este apartado, se define la pregunta de investigación de forma que se focalice el área de interés del trabajo y queden definidos tanto el problema a tratar como sus principales características.

A. Objeto de la pregunta.

En esta revisión sistemática se pretende localizar trabajos centrados en el desarrollo de modelos y metodologías de análisis de riesgos, con el objetivo de que puedan ser aplicadas en PYMES y puedan adaptarse a cubrir riesgos asociativos y jerárquicos.

B. Premisa de la Revisión Sistemática.

Podemos definir la pregunta de investigación de este trabajo, por tanto, de la siguiente forma:

¿Qué trabajos se han llevado a cabo para desarrollar sistemas de análisis de riesgos teniendo en cuenta riesgos jerárquicos, asociativos y aplicación en PYMES?

Las palabras y conceptos relacionados que se utilizaron para formular esta pregunta y que fueron utilizados durante la ejecución de la revisión son las siguientes:

Risk analysis: risk analysis model, risk analysis methodology
Risk management: risk management model, risk management methodology
Risks: Associative risks, hierarchical risks
Small & Medium-Sized Business: SMB, SME, PYME

En el contexto de la revisión sistemática planificada se van

a observar las propuestas existentes sobre modelos y metodologías de análisis de riesgos, haciendo especial hincapié en aquellas orientadas a trabajo con riesgos asociativos, riesgos jerárquicos y/u orientadas a PYMES, extrayendo las más importantes y procediendo a un posterior análisis y comparación de las mismas. La población a analizar se compone de las publicaciones presentes en los repositorios de las fuentes de datos seleccionadas que estén relacionadas con el objetivo de esta revisión.

Los resultados esperados de esta revisión son conocer las propuestas existentes en cuanto a análisis de riesgos asociativos y jerárquicos con orientación a PYMES, para posteriormente analizarlas y conocer qué comparten y en qué difieren, además de identificar necesidades de investigación. Las principales áreas de aplicación que se verán beneficiadas por los resultados de esta revisión sistemática son las relacionadas con la Seguridad de la Información, en especial la Gestión de la Seguridad (concretamente los análisis de riesgos), así como los expertos en seguridad. A tal fin, se proveerá un marco comparativo que permita posicionar adecuadamente las nuevas actividades de investigación en análisis de riesgos.

III. MÉTODO DE LA REVISIÓN

El método de revisión se basa en el protocolo de investigación. En esta etapa definimos la estrategia de búsqueda, qué fuentes se utilizarán para identificar los estudios primarios, si hubo algunas restricciones, cuáles son los criterios de inclusión y exclusión, qué criterios se utilizarán para evaluar la calidad de los estudios primarios y cómo se extraerán y sintetizarán los datos de los estudios.

A. Selección de fuentes.

El objetivo de esta fase es seleccionar las fuentes que se usarán para realizar la ejecución de la búsqueda de estudios primarios.

El criterio para la selección de las fuentes de búsqueda será la posibilidad de consultar los documentos en Internet o en la biblioteca digital de la Universidad de Castilla-La Mancha, que cuenta con libros electrónicos así como con acceso a las bibliotecas digitales de ACM, IEEE, Science@-Direct o Elsevier, entre otros; la inclusión motores de búsqueda que permitan consultas avanzadas y búsqueda por palabras clave; además, editoriales, libros, revistas y conferencias recomendadas por expertos en la materia (como los miembros de RETISTRUST1, una Red española de expertos en Seguridad de la Información).

La búsqueda de estudios primarios se llevará a cabo utilizando motores de búsqueda en web, bases de datos electrónicas y búsquedas manuales, tales como búsquedas en una revista/conferencia/libro/publicación específica o en publicaciones de investigación recomendadas por expertos en la materia.

Finalmente, la lista de fuentes inicial obtenida sobre la cual se ejecutará la revisión sistemática es la siguiente: ACM digital library, IEEE digital library, Science@Direct, Google

Scholar, SREIS symposium, ESORICS symposium, REFSQ conference, IEEE International Requirements Engineering Conference, ICSE conference, COMPSAC conference, DEXA conference, WOSIS workshop, ICCSA conference, Requirements Engineering Journal, Computer Standards & Interfaces Journal, Computers & Security.

B. Selección de estudios.

Una vez que se han sido definidas las fuentes, es necesario describir el proceso y el criterio que vamos a seguir en la ejecución de la revisión para la selección y evaluación de los estudios.

En primer lugar, se combinaron las palabras clave seleccionadas con conectores AND y OR para obtener la cadena de búsqueda, como se muestra a continuación:

methodology OR model
AND
associative OR hierarchical
AND
"risk analysis" OR "risk management" OR "risk
assessment"
AND
SMB OR SME OR PYME

El procedimiento para la selección de estudios empleado comienza con la adaptación de la cadena de búsqueda al motor de búsqueda de la fuente y la ejecución de la consulta, limitando la búsqueda a trabajos publicados en los últimos 5 años. Los criterios de inclusión y exclusión deberían basarse en la Pregunta de investigación. El criterio de inclusión actúa sobre los resultados obtenidos al ejecutar la búsqueda sobre la fuente, permitiéndonos realizar una primera selección de documentos que serán considerados en el contexto de la revisión como candidatos a convertirse en estudios primarios. Como criterio de inclusión se realiza principalmente un análisis sobre el título, las palabras claves y el abstract de cada documento, de forma que podemos ver en una primera instancia cómo están relacionadas estas palabras y por qué ha sido seleccionado el estudio. Con este criterio se localizan y eliminan la mayor parte de los resultados obtenidos que no realizan aportaciones sobre análisis de riesgos en el campo de los Sistemas de Información.

El criterio de exclusión actúa sobre el subconjunto de estudios relevantes obtenidos y nos permite obtener el conjunto de estudios primarios. En esta fase nos centramos principalmente en la lectura y análisis del abstract del documento y sus conclusiones, teniendo en algunos casos que profundizar en el mismo y realizar una lectura más detallada sobre otras partes del documento, de cara a ver en más detalle de qué trata cada estudio, ver la relación real que presenta con los objetivos buscados y, si es verdaderamente relevante para la revisión, seleccionarlo como estudio primario.

C. Ejecución de la selección.

En este punto, se ejecuta la búsqueda en cada una de las fuentes seleccionadas con el fin de obtener una lista inicial de los estudios para la posterior evaluación aplicando todos los criterios y procedimientos especificados.

Los procedimientos para la selección de los estudios se aplican a todos los artículos obtenidos a fin de verificar si los estudios se ajustan a los criterios de inclusión y exclusión. Los estudios obtenidos, que corresponden exactamente con todos los criterios de inclusión y exclusión definidos previamente, se detallan en la siguiente sección.

IV. EXTRACCIÓN DE LA INFORMACIÓN.

La información extraída de los estudios debe contener las técnicas, métodos, procesos, medidas, estrategias o cualquier tipo de iniciativa para la adaptación del análisis, gestión o evaluación de riesgos a un alcance abordable por las PYMES, o manejar riesgos asociativos o jerárquicos.

Los formularios de información definidos para esta revisión sistemática contienen la identificación del estudio, la metodología o modelo del estudio, los resultados del estudio, los problemas del estudio y nuestras impresiones generales al respecto.

A continuación se ofrece una breve reseña de cada uno de los estudios seleccionados mostrados en la sección anterior, de acuerdo con la información extraída obtenida a través de los formularios de información.

4.1. Feng, Nan et al. "A security risk analysis model for information systems: Causal relationships of risk factors and vulnerability propagation analysis" [21].

Los autores presentan un modelo de análisis de riesgos de seguridad con el objetivo de identificar las relaciones causales entre los factores de riesgo y analizar la complejidad y la incertidumbre de la propagación de las vulnerabilidades. Se basan en que en los sistemas de información, los riesgos de seguridad son causados por diversos factores internos y externos interrelacionados. De esta forma, una vulnerabilidad de seguridad también podría propagarse y escalar a través de las cadenas causales de los factores de riesgo a través de diferentes vías de acceso.

Los autores desarrollan una red Bayesiana para definir simultáneamente los factores de riesgo y sus relaciones causales basadas en el conocimiento a partir de los casos observados y de los expertos en el dominio.

Se trata de un estudio teórico, sin contrastar resultados con aplicación de la propuesta en casos prácticos.

4.2. Webb, J. et al. "A situation awareness model for information security risk management" [22].

Los autores proponen un modelo de análisis de riesgos de seguridad de la información consciente de la situación (SA-ISRM) para complementar el proceso de gestión del riesgo de seguridad de la información. Su objetivo es paliar las

deficiencias en la práctica de la evaluación de riesgos de seguridad de información que inevitablemente conducen a una mala toma de decisiones y estrategias inadecuadas o inapropiadas de seguridad.

De esta forma, el modelo propuesto busca responder a dichas deficiencias a través de la recogida, análisis y comunicación de la información relacionada con los riesgos de la empresa en su totalidad.

El modelo ha sido refinado y perfeccionado mediante un caso de estudio en la empresa de inteligencia de seguridad nacional de Estados Unidos.

4.3. Yongli Tang et al. "Information Security Risk Assessment Method based on Cloud Model" [23].

Los autores proponen una metodología de evaluación del riesgo en Sistemas de Información basándose en la construcción de un modelo basado en la Nube (Cloud). De esta forma, su objetivo es utilizar el "Modelo Cloud" para reducir la incertidumbre en la cuantificación de resultados del análisis de riesgos. Este modelo, por su propia naturaleza, permite tener en cuenta factores asociativos a la hora de realizar la evaluación de riesgos, empleando para ello técnicas difusas (fuzzy).

Se trata de un estudio teórico, sin contrastar resultados con aplicación de la propuesta en casos prácticos.

4.4. Vicente, E. et al. "Risk analysis in information systems: A fuzzification of the MAGERIT methodology" [24].

Los autores presentan una extensión de la metodología MAGERIT basada en modelos computacionales fuzzy (difusos) con el objetivo de reducir el grado de incertidumbre en las técnicas de medición de las metodologías tradicionales.

De esta forma, presentan una escala de términos lingüísticos para representar los valores de medición, sus dependencias y frecuencias y la degradación de los activos en entornos de Sistemas de Información.

Estas técnicas se aplican teniendo en cuenta también que la relación de activos de SI puede ser tanto interna como depender de terceras partes, lo que apoya la necesidad de trabajar con factores asociativos de cara a la evaluación y gestión de riesgos.

Se trata de un estudio teórico, sin contrastar resultados con aplicación de la propuesta en casos prácticos.

4.5. Saptarshi, M. et al. "Risk analysis using FMEA: Fuzzy similarity value and possibility theory based approach" [25].

Los autores proponen una metodología que incorpore a los procesos de análisis de riesgos las técnicas de FMEA (Failure Mode and Effect Analysis), particularmente las aproximaciones basadas en reglas y técnicas fuzzy.

El objetivo principal es utilizar estas técnicas para reducir la arbitrariedad y, con ello, la incertidumbre en el análisis de

riesgos, integrando conceptos de similitud de valores de medición de los número difusos y teorías de posibilidad.

Se trata de un estudio teórico, sin contrastar resultados con aplicación de la propuesta en casos prácticos.

4.6. Abdel-Basset, Mohamed et al. "A framework for risk assessment, management and evaluation: Economic tool for quantifying risks in supply chain" [26].

Los autores proponen un framework de gestión y evaluación de riesgos basado en la aplicación de técnicas neutrosóficas. Aunque se aplica al riesgo en el ámbito de una cadena de montaje, es interesante estudiar su posible adaptación a los Sistemas de Información.

Lo interesante de esta propuesta es la introducción de la incertidumbre dentro del análisis de riesgos, así como la posibilidad de obtener valores cuantitativos de riesgo dentro de un escenario afectado por sucesos imprevistos y al que pueden afectar factores como la subjetividad, la incertidumbre o la vaguedad a la hora de obtener valores de riesgo cuantitativos y objetivos que puedan orientar procesos de tomas de decisión. Los autores emplean el Proceso Analítico Jerárquico Neutrosófico (N-AHP) para analizar los factores de riesgo identificados.

Los autores han definido un caso de estudio para aplicar y refinar este framework.

4.7. Sicari, S. et al. "A risk assessment methodology for the Internet of Things" [27].

Los autores proponen una Metodología de análisis y gestión de riesgos con aplicación sobre entornos IoT. El método propuesto (tanto cualitativo como cuantitativo) se basa la construcción de un árbol de ataques adaptado a cada escenario y en un criterio denominado valor de explotabilidad. Inicialmente, la evaluación de este valor se obtiene de forma cualitativa, considerando los niveles de dificultad de realizar un ataque contra el sistema. Estos niveles cualitativos se traducen posteriormente en valores cuantitativos concretos. El valor de explotabilidad general del sistema se calcula finalmente sobre la base de un grafo de dependencia entre las vulnerabilidades identificadas.

El procedimiento propuesto es fundamentalmente teórico. Se aplica en un caso práctico, pero es demasiado global y sin detallar demasiado los procesos llevados a cabo para obtener los resultados. Además, requiere un alto grado de conocimiento experto para su mantenimiento, y se centra sobre todo en riesgo de ataque sobre componentes físicos, siendo demasiado específico.

4.8. Staalduinen, M.A. et al. "Functional quantitative security risk analysis (QSRA) to assist in protecting critical process infrastructure" [28].

Los autores proponen una metodología cuantitativa de evaluación de riesgos de seguridad orientada a infraestructuras críticas. Se parte de un enfoque orientado a la evaluación

concurrente de amenazas y vulnerabilidades y se introduce un modelo de riesgo “Bow Tie” mapeado en un modelo de Red Bayesiana que permite diferentes supuestos lógicos. Finalmente, se integran las probabilidades de riesgo/vulnerabilidad con valores de pérdida potencial para cuantificar el riesgo.

La importancia de procesos de análisis de riesgos adaptados a infraestructuras críticas siguiendo un modelo “Bow Tie” también se presenta por Abdo, H. et al. en “A safety/security risk analysis approach of Industrial Control Systems: A cyber bowtie – combining new version of attack tree with bowtie analysis” [29].

Aunque ambos puntos de partida se centran sobre instalaciones químicas (sobre todo de cara a preparar casos de estudio reales), las metodologías son adaptables a cualquier infraestructura crítica configurando y personalizando sus elementos, aunque es necesario el conocimiento experto para llevarla a cabo.

Los autores han definido un caso de estudio para aplicar y refinar esta metodología.

4.9. Khan, F. et al. “Dynamic risk management: a contemporary approach to process safety management” [30].

Los autores proponen un framework para la gestión dinámica de riesgos, cuya piedra angular es un proceso de evaluación dinámica de riesgos basado en una estrategia Plan-Do-Check-Act (PDCA). De esta forma, se define una evaluación inicial de riesgos, tras la que comienza un ciclo PDCA de evaluación continua.

El framework propuesto está aún en una fase muy inicial, pero lo más interesante de esta propuesta es la creciente importancia del concepto de dinamismo dentro de los procesos de evaluación de riesgos.

Se trata de un estudio teórico, sin contrastar resultados con aplicación de la propuesta en casos prácticos.

4.10. Munodawafa, F. et al. “Security risk assessment within hybrid data centers: A case study of delay sensitive applications” [31].

Los autores presentan un estudio discursivo sobre la necesidad de contar con procesos de análisis y gestión de riesgos específicamente en el ámbito de los Centros de datos. No se llega a proponer ni definir un mecanismo concreto pero pone sobre la mesa conceptos relevantes como la necesidad de incluir en la seguridad de estos Centros de datos los riesgos a los que están sometidos no sólo los servidores físicos sino también los virtuales. Este nuevo escenario entronca con las nuevas necesidades en el área del Cloud Computing, con la necesidad de convivencia de sistemas físicos clásicos con sistemas virtuales, así como en los riesgos asociativos derivados de la virtualización.

El estudio presenta una selección inicial de Riesgos y Vulnerabilidades centrados en Data centers, incluyendo algunos específicos para servidores virtuales. También se presenta un caso de estudio centrado específicamente en la

evaluación de aspectos de Disponibilidad sobre un servicio de VoIP, si bien es aún muy esquemático y sin demasiado detalle.

4.11. Panchal, D. et al. “A new fuzzy methodology-based structured framework for RAM and risk analysis” [32].

Los autores proponen un framework para la realización de análisis de riesgos que sustenta sus procesos de análisis de riesgos en técnicas de FMEA (Failure Mode and Effect Analysis), particularmente las aproximaciones basadas en reglas y técnicas fuzzy.

El objetivo principal es utilizar estas técnicas para reducir la arbitrariedad y, con ello, la incertidumbre en el análisis de riesgos. De esta forma se utiliza un enfoque Fuzzy Lambda – Tau (FLT) para calcular los parámetros de confiabilidad, disponibilidad y mantenibilidad (RAM) del sistema.

El estudio se centra en el ámbito de una planta de proceso químico, aunque es lo suficientemente genérico como para adaptarse a cualquier tipo de Sistema de Información. En todo caso, refuerza la creciente importancia de procesos de análisis de riesgo específicos para Infraestructuras críticas.

Se trata de un estudio teórico, sin contrastar resultados con aplicación de la propuesta en casos prácticos.

4.12. Sangaiah, A.K. et al. “Towards an efficient risk assessment in software projects–Fuzzy reinforcement paradigm” [33].

Los autores proponen un acercamiento basado en técnicas fuzzy como base para el futuro desarrollo de un framework de evaluación de riesgos que permita manejar la incertidumbre y evaluar de forma eficiente los riesgos en el ámbito del desarrollo de proyectos software, de forma que puedan guiar un proceso de tomas de decisiones eficiente a lo largo del Ciclo de vida del proyecto.

Se trata de un estudio teórico, sin contrastar resultados con aplicación de la propuesta en casos prácticos.

4.13. Wangen, G. et al. “A framework for estimating information security risk assessment method completeness” [34].

Los autores proponen un framework llamado CURF (Core Unified Risk Framework) cuyo objetivo es comparar métodos de evaluación de riesgos de sistemas de información.

La propuesta es interesante por plantear la necesidad de que este framework sea dinámico, permitiendo adaptarse a nuevas características y tareas de los métodos revisados. Además, entre sus criterios de comparación de métodos de análisis de riesgo incluye como factores clave que se adapte a Cloud Computing y que tenga en cuenta la Reutilización del conocimiento.

4.14. Zhang, H. et al. “An Integrated Approach to Risk Assessment for Special Line Shunting Via Fuzzy Theory” [35].

Los autores presentan un proceso de evaluación de riesgos basado en técnicas fuzzy con el objetivo principal de obtener valores de riesgo fiables en entornos sujetos a factores ambientales que condicen a obtener resultados de riesgo incompletos o involucran altos niveles de incertidumbre.

Aunque el estudio se aplica específicamente al ámbito ferroviario, sus conceptos son fácilmente extrapolables al campo de las TI donde la incertidumbre es también un factor clave en los procesos de Análisis de riesgos. De esta forma, se redunda en la importancia de la reducción de incertidumbre de cara a obtener resultados fiables, así como en otros conceptos muy interesantes como son la utilización tanto de técnicas cualitativas como cuantitativas.

El estudio también incide en la necesidad de tener en cuenta las relaciones jerárquicas y define un caso de estudio para aplicar este proceso.

V. ANÁLISIS DE RESULTADOS.

Los resultados de la revisión sistemática se muestran en la siguiente tabla, que resume la cantidad de estudios por iniciativa:

Tipo de iniciativa	Nº de estudios	Iniciativas
Proceso	1	4.14
Framework	5	4.6, 4.9, 4.11, 4.12, 4.13
Modelo	2	4.1, 4.2
Metodología	5	4.3, 4.4, 4.5, 4.7, 4.8
Otros	1	4.10
Total	14	-

Tabla 1: Resultados por iniciativa

Como podemos ver en la tabla anterior, hay muchos nuevos frameworks, procesos, modelos y metodologías que intentan facilitar la gestión, evaluación y/o el análisis de riesgos teniendo en cuenta factores como la flexibilidad o simplicidad de su aplicación (necesaria para poder aplicarlos al ámbito de la PYME), o considerando la importancia de gestionar los riesgos jerárquicos y asociativos, imprescindible para Cloud Computing o IoT, por ejemplo. También se empieza a ver la necesidad de controlar el riesgo en sistemas asociados de forma específica a Infraestructuras críticas.

Sin embargo, muy pocos trabajos describen casos de estudio complejos que muestren la posibilidad y los beneficios obtenidos de aplicar el modelo o metodología propuestos en la práctica.

Por otra parte, como se puede ver en la Tabla 2, después de nuestro análisis hemos llegado a la conclusión de que cada una de las iniciativas seleccionadas nos ofrece aspectos muy importantes que tienen que ver con los requisitos de análisis de riesgos en Sistemas de Información. Estas son

características que pueden ser utilizadas como base para el desarrollo de una metodología que incluya todas las características deseadas.

Iniciativa	Ámbito	Técnica/Modelo base	Principales contribuciones
Feng, Nan	Global	Redes Bayesianas	- Análisis de la incertidumbre de propagación de debilidades
Webb, J.	Global	-	- Procesos de toma de decisiones - Mejora de los procesos de recogida, análisis y comunicación de información relativa a riesgos - Caso de estudio
Yongli Tang	Global	Modelo Cloud Técnicas difusas	- Modelo basado en la Nube - Gestión de riesgos asociativos
Vicente, E.	Global	Técnicas difusas	- Reducción del grado de incertidumbre - Importancia del entorno y de terceros en la evaluación de riesgos
Saptarshi, M.	Global	FMEA	- Reducción del grado de incertidumbre
Abdel-Basset, Mohamed	-	N-AHP	- Gestión de la incertidumbre - Uso de AHP para analizar factores de riesgo e impacto - Caso de estudio
Sicari, S.	IoT	-	- Valor de explotabilidad de vulnerabilidades - Grafos de vulnerabilidad
Staalduinen, M.A. Abdo, H.	Infraestructuras Críticas	- Modelos Bow Tie - Redes Bayesianas	- Análisis concurrente de riesgos y vulnerabilidades - Caso de estudio
Khan, F.	Global	-	- Evaluación dinámica del riesgo
Munodawafa, F.	Data Centers	-	- Análisis de riesgos en virtualización - Riesgos en Data centers fuera de la infraestructura del SI pero que forman parte de ella
Panchal, D.	Infraestructuras Críticas	FMEA	- Reducción del grado de incertidumbre
Sangaiah, A.K.	Desarrollo de Software	Técnicas difusas	- Reducción del grado de incertidumbre
Wangen, G.	Global	-	- Reutilización del conocimiento - Importancia de la Evaluación dinámica del riesgo y los entornos en la Nube
Zhang, H.	-	Técnicas difusas	- Reducción del grado de incertidumbre - Factores jerárquicos para valoración de probabilidad e impacto de los riesgos

Tabla 2: Principales contribuciones de las propuestas seleccionadas

En la Tabla 3 se puede ver una comparativa de las diferentes propuestas analizadas, comparadas con la propuesta futura que pretende abordarse. Se considera que los aspectos valorados se pueden cumplir de forma total, parcialmente o no haber sido abordados en el modelo. A continuación, se describe cada uno de los aspectos analizados:

- **Ámbito de aplicación:** Si el modelo se aplica de forma global a la seguridad los Sistemas de Información de una compañía, o sólo a un subconjunto de ellos.
- **Métricas:** La guía incluye mecanismos de medición de los criterios de riesgo claros, detallando información sobre su aplicación y evaluación.
- **Técnicas cualitativas:** El modelo incluye técnicas cualitativas de medición.
- **Técnicas cuantitativas:** El modelo incluye técnicas cuantitativas de medición.
- **Asociativo:** El modelo tiene en cuenta la distribución del riesgo (por ejemplo, funciones derivadas a terceros, o realizadas por la empresa en colaboración con otras empresas) y la interrelación de la empresa con el entorno.
- **Jerárquico:** El modelo tiene en cuenta la relación jerárquica entre compañías relacionadas. (Por ejemplo, el esquema Matriz – Filiales).
- **Orientado a PYMES:** El modelo ha sido desarrollado pensando en la casuística especial de las PYMES.
- **Reutiliza el conocimiento:** La guía adquiere conocimiento de las implantaciones y de la información recogida durante su utilización, de forma que este conocimiento pueda ser reutilizado para facilitar posteriores implantaciones.
- **Dispone de herramienta software:** El modelo dispone de una herramienta que lo soporte.
- **Casos prácticos:** El modelo ha sido desarrollado y refinado a partir de casos prácticos.
- **Cloud Computing:** El modelo tiene en cuenta la aplicación en entornos de Cloud Computing.

Estas características deseables para un modelo de análisis y gestión de riesgos asociativos y jerárquicos para PYMES se han obtenido a través de la aplicación del "método de investigación-acción" a casos reales. Se considera que cada uno de estos aspectos puede ser totalmente cumplido (S), parcialmente cumplido (P) o no tenido en cuenta por el modelo (N).

Iniciativa	Ámbito Global	Métricas	Técnicas Cualitativas	Técnicas Cuantitativas	Asociativo	Jerárquico	Orientado PYMES	Reutilización Conocimiento	Herramienta Software	Casos Prácticos	Cloud Computing
Feng, Nan	S	S	N	S	S	N	N	N	N	N	N
Webb, J.	S	N	N	N	S	N	N	N	N	S	N
Yongli Tang	S	S	S	S	S	N	N	N	N	N	S
Vicente, E.	S	S	S	N	S	N	N	N	N	N	P
Saptarshi, M.	S	S	S	N	S	N	N	N	N	N	N

Iniciativa	Ámbito Global	Métricas	Técnicas Cualitativas	Técnicas Cuantitativas	Asociativo	Jerárquico	Orientado PYMES	Reutilización Conocimiento	Herramienta Software	Casos Prácticos	Cloud Computing
Abdel-Basset, Mohamed	N	N	N	S	P	N	N	N	N	S	N
Sicari, S.	N	P	S	S	N	N	N	N	N	S	N
Staalduinen, M.A.	N	N	N	S	N	N	N	N	N	S	N
Abdo, H.											
Khan, F.	S	N	N	N	N	N	N	P	N	N	N
Munodawafa, F.	N	N	N	N	P	N	N	N	N	P	S
Panchal, D.	S	P	S	N	N	N	N	N	N	N	N
Sangaiah, A.K.	N	S	S	N	N	N	N	N	N	N	N
Wangen, G.	S	N	N	N	S	N	N	S	N	N	N
Zhang, H.	N	S	S	S	N	S	N	N	N	N	N
MARISMA	S	S	S	S	S	S	S	S	S	S	S

Tabla 3: Comparativa de las propuestas seleccionadas

Se puede ver cómo ninguna de las propuestas estudiadas posee las características requeridas por las PYMES:

- No están pensadas para su aplicación en empresas de pequeño tamaño y, por tanto, con escasos recursos humanos y económicos.
- La mayoría se centran sólo en el análisis de riesgos de una parte del Sistema de Información, y casi ninguna de ellas aborda desde un punto de vista global la implantación de estos sistemas, lo que obligaría a las compañías a tener que adquirir, implementar, gestionar y mantener varias metodologías, modelos y herramientas para gestionar de forma integral los riesgos. Adicionalmente, las pocas aplicaciones que han intentado abordar todo el Sistema de Información requieren de una gestión compleja y de un mantenimiento costoso, con una necesidad importante de conocimiento experto para poder mantener el sistema de gestión y evaluación de riesgos, lo que hace que no sean adecuadas para las PYMES.
- La mayoría de las propuestas no tienen en cuenta la necesidad de contemplar riesgos jerárquicos y asociativos y, las que lo hacen, lo realizan desde un punto de vista teórico, sin establecer mecanismos concretos y basados en casos prácticos para gestionar este tipo de riesgos.

Por lo tanto, es relevante realizar un nuevo modelo que permita incluir todas esas características, incluyendo la automatización de las métricas para reducir los costes de mantenimiento del sistema.

VI. CONCLUSIONES.

En este artículo se ha realizado una revisión sistemática de los diferentes modelos y metodologías para el análisis y gestión de riesgos, con el objetivo de estudiar las propuestas centradas en riesgos asociativos y jerárquicos, que sean válidas para las PYMES.

Como resultado de esta revisión se ha podido establecer la importancia que tiene la gestión y el análisis de los riesgos sobre la seguridad de los Sistemas de Información en el desempeño y evolución sostenible de las empresas, ya que constituye un requisito básico para alcanzar la misión y los objetivos organizacionales en un entorno altamente competitivo.

En numerosas fuentes bibliográficas se detecta y resalta la dificultad que supone para las PYMES la utilización de las metodologías y modelos de análisis de riesgos tradicionales, que han sido concebidos para grandes empresas, siendo la aplicación de este tipo de metodologías y modelos difícil y costosa para las PYMES [36-40].

El problema principal de todos los modelos de análisis y gestión riesgos existentes es que no están teniendo éxito a la hora de implantarse en PYMES, debido principalmente a que:

- Unos fueron desarrollados pensando en organizaciones grandes (Grandes estándares como CRAMM, ISO/IEC 27005, MAGERIT, OCTAVE, NIST SP 800-39, Mehari, COBIT o ERMF) y en las estructuras organizativas asociadas a éstas.
- Otros (Khan [30]) han intentado simplificar el modelo para que pudiera ser apto para compañías con recursos limitados, pero son modelos incompletos que sólo afrontan parte del problema, o intentan aportar unas guías básicas de los pasos a realizar, pero sin entrar en cómo evaluar y gestionar realmente los riesgos de una forma en la que el propio personal técnico de la empresa se pueda involucrar. Además, la mayoría son modelos teóricos y están todavía en desarrollo.
- La mayoría de las propuestas no tienen en cuenta la necesidad de contemplar riesgos jerárquicos y asociativos, factores cruciales en la estructura y funcionamiento actual de las empresas (en el que cada vez tiene más peso el uso de sistemas en Cloud), sobre todo de las PYMES.
- Aunque está creciendo el número de propuestas que inciden en la necesidad de tener en cuenta factores como el dinamismo, la reutilización del conocimiento y la reducción de la incertidumbre a la hora de realizar un análisis de riesgos, son muy pocas las traslaciones prácticas de dichas propuestas a casos reales y se encuentran en fases muy iniciales.

De esta forma, se puede concluir que es relevante realizar un nuevo modelo que permita incluir todas las características citadas como deseables de cara a su implantación en todo tipo de compañías, y en especial para el caso de las PYMES.

Todos los estándares y propuestas para la evaluación y gestión de riesgos estudiados son muy importantes, y sus

aportaciones serán tenidas en cuenta para el desarrollo de una metodología que incluya todas las características deseadas.

AGRADECIMIENTOS

Esta investigación ha sido co-financiada por los proyectos *GENESIS - Security Government of Big Data and Cyber Physics Systems ((SBPLY/17/180501/000202)* financiado por el “JCCM- Consejería de Educación, Cultura y Deportes, y Fondos FEDER”, del proyecto *ECLIPSE – Enhancing Data Quality and Security for Improving Business Processes and Strategic Decisions in Cyber Physical Systems (RTI2018-094283-B-C31)* financiado por la “Ministerio Economía, Industria y Competitividad y fondos FEDER”, y ha contado con el apoyo de las empresas Marisma Shield S.L (www.emarisma.com) y Sicaman Nuevas Tecnologías (www.sicaman-nt.com).

REFERENCIAS

- [1] Kluge, D. *Formal Information Security Standards in German Medium Enterprises*. in *CONISAR: The Conference on Information Systems Applied Research*. 2008.
- [2] Dhillon, G. and J. Backhouse, *Information System Security Management in the New Millennium*. Communications of the ACM, 2000. **43**(7): p. 125-128.
- [3] Vivas, T., A. Zambrano, and M. Huerta. *Mechanisms of security based on digital certificates applied in a telemedicine network*. in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2008. IEEE.
- [4] Barlette, Y. and V. Vladislav. *Exploring the Suitability of IS Security Management Standards for SMEs*. in *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*. 2008. Waikoloa, HI, USA.
- [5] Wiander, T. and J. Holappa, *Theoretical Framework of ISO 17799 Compliant. Information Security Management System Using Novel ASD Method.*, in *Technical Report*, V.T.R.C.o. Finland, Editor 2006.
- [6] Huerta, M., et al. *Implementation of a open source security software platform in a telemedicine network*. in *Proceedings of the 9th WSEAS international conference on Advances in e-activities, information security and privacy*. 2010. World Scientific and Engineering Academy and Society (WSEAS).
- [7] Pirrone, J. and M. Huerta. *Security Mechanism for Medical Record Exchange Using Hippocratic Protocol*. in *World Congress on Medical Physics and Biomedical Engineering 2018*. 2019. Springer.
- [8] Wiander, T. *Implementing the ISO/IEC 17799 standard in practice – experiences on audit phases*. in *AISC '08: Proceedings of the sixth Australasian conference on Information security*. 2008. Wollongong, Australia.
- [9] Huerta, M., et al. *Design of a building security system in a university campus using RFID technology*. in *2017 IEEE 37th Central America and Panama Convention (CONCAPAN XXXVII)*. 2017.
- [10] Michalson, L., *Information security and the law: threats and how to manage them*. Convergence, 2003. **4**(3): p. 34-38.
- [11] Volonino, L. and S. Robinson. *Principles and Practice of Information Security*. in 1 edition, Anderson, Natalie E. 2004. New Jersey.
- [12] Spinellis, D. and D. Gritzalis. *Information Security Best Practise Dissemination: The ISA-EUNET Approach*. in *WISE 1: First World Conference on Information Security Education*. 1999.
- [13] Siegel, C.A., T.R. Sagalow, and P. Serritella, *Cyber-Risk Management: Technical and Insurance Controls for Enterprise-Level Security*. Security Management Practices, 2002. **sept/oct**: p. 33-49.
- [14] Garigue, R. and M. Stefani, *Information Security Governance Reporting*. Information Systems Security, 2003. **sept/oct**: p. 36-40.
- [15] Mercuri, R.T., *Analyzing security costs*. Communications of the ACM, 2003. **46**: p. 15-18.

- [16] Zissis, D. and D. Lekkas, *Addressing cloud computing security issues*. Future Generation Computer Systems, 2012. **28**(3): p. 583-592.
- [17] Brereton, P., et al., *Lessons from applying the systematic literature review process within the software engineering domain*. Journal of Systems and Software, 2007. **80**(4): p. 571-583.
- [18] Kitchenham, B., *Procedures for performing systematic reviews*. Keele, UK, Keele University, 2004. **33**(2004): p. 1-26.
- [19] Kitchenham, B. and S. Charters, *Guidelines for performing systematic literature reviews in software engineering version 2.3*. Engineering, 2007. **45**(4ve): p. 1051.
- [20] Biolchini, J., et al., *Systematic review in software engineering*. System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES, 2005. **679**(05): p. 45.
- [21] Feng, N., H.J. Wang, and M. Li, *A security risk analysis model for information systems: Causal relationships of risk factors and vulnerability propagation analysis*. Information Sciences, 2014. **256**(0): p. 57-73.
- [22] Webb, J., et al., *A situation awareness model for information security risk management*. Computers & Security, 2014. **44**(0): p. 1-15.
- [23] Yongli, T., et al. *Information security risk assessment method based on cloud model*. in *Irish Signals & Systems Conference 2014 and 2014 China-Ireland International Conf. on Infor. and Comm. Technologies (ISSC 2014/CIICT 2014)*. 25th IET. 2014.
- [24] Vicente, E., A. Mateos, and A. Jiménez-Martin, *Risk analysis in information systems: A fuzzification of the MAGERIT methodology*. Knowledge-Based Systems, 2014. **66**(0): p. 1-12.
- [25] Mandal, S. and J. Maiti, *Risk analysis using FMEA: Fuzzy similarity value and possibility theory based approach*. Expert Systems with Applications, 2014. **41**(7): p. 3527-3537.
- [26] Abdel-Basset, M., et al., *A framework for risk assessment, management and evaluation: Economic tool for quantifying risks in supply chain*. Future Generation Computer Systems, 2019. **90**: p. 489-502.
- [27] Sicari, S., et al., *A risk assessment methodology for the Internet of Things*. Computer Communications, 2018. **129**: p. 67-79.
- [28] van Staalduinen, M.A., et al., *Functional quantitative security risk analysis (QSR4) to assist in protecting critical process infrastructure*. Reliability Engineering & System Safety, 2017. **157**: p. 23-34.
- [29] Abdo, H., et al., *A safety/security risk analysis approach of Industrial Control Systems: A cyber bowtie—combining new version of attack tree with bowtie analysis*. Computers & Security, 2018. **72**: p. 175-195.
- [30] Khan, F., et al., *Dynamic risk management: a contemporary approach to process safety management*. Current opinion in chemical engineering, 2016. **14**: p. 9-17.
- [31] Munodawafa, F. and A.I. Awad, *Security risk assessment within hybrid data centers: A case study of delay sensitive applications*. Journal of Information Security and Applications, 2018. **43**: p. 61-72.
- [32] Panchal, D., et al., *A new fuzzy methodology-based structured framework for RAM and risk analysis*. Applied Soft Computing, 2019. **74**: p. 242-254.
- [33] Sangaiah, A.K., et al., *Towards an efficient risk assessment in software projects—Fuzzy reinforcement paradigm*. Computers & Electrical Engineering, 2018. **71**: p. 833-846.
- [34] Wangen, G., C. Hallstensen, and E. Snekenes, *A framework for estimating information security risk assessment method completeness*. International Journal of Information Security, 2018. **17**(6): p. 681-699.
- [35] Zhang, H. and Q. Sun, *An Integrated Approach to Risk Assessment for Special Line Shunting Via Fuzzy Theory*. Symmetry, 2018. **10**(11): p. 599.
- [36] Batista, J. and A. Figueiredo, *SPI in very small team: a case with CMM*. Software Process Impr. and Practice, 2000. **5**(4): p. 243-250.
- [37] Hareton, L. and Y. Terence, *A Process Framework for Small Projects*. Software Process Improvement and Practice, 2001. **6**: p. 67-83.
- [38] Tuffley, A., B. Grove, and M. G., *SPICE For Small Organisations*. Software Process Improvement and Practice, 2004. **9**: p. 23-31.
- [39] Calvo-Manzano, J.A., et al., *Experiences in the Application of Software Process Improvement in SMES*. Software Quality Journal, 2004. **10**(3): p. 261-273.
- [40] Mekelburg, D., *Sustaining Best Practices: How Real-World Software Organizations Improve Quality Processes*. Software Quality Professional, 2005. **7**(3): p. 4-13.

Antonio Santos-Olmo is MsC in Computer Science and is an Assistant Professor at the Escuela Superior de Informática of the Universidad de Castilla- La Mancha in Ciudad Real (Spain) (Computer Science Department, University of Castilla La Mancha, Ciudad Real, Spain), MSc in Information Systems Audit from the Polytechnic University of Madrid, and Certified Information System Auditor by ISACA. He is the Director of Software Factory departments of the company Sicaman Nuevas Tecnologías S.L. His research activities are management security system, security metrics, data mining, data cleaning, and business intelligence. He participates in the GSyA research group of the Department of Computer Science at the University of Castilla- LaMancha, in Ciudad Real (Spain).

Luis Enrique Sánchez is PhD and MsC in Computer Science and is a Professor at the Universidad de Castilla-la Mancha (Ciudad Real, Spain), MSc in Information Systems Audit from the Polytechnic University of Madrid, and Certified Information System Auditor by ISACA. He is the Director of Professional Services and R&D departments of the company Sicaman Nuevas Tecnologías S.L. COIICLM board or committee member and responsible for the professional services committee. His research activities are management security system, security metrics, data mining, data cleaning, and business intelligence. He participates in the GSyA research group of the Department of Computer Science at the University of Castilla- LaMancha, in Ciudad Real (Spain).

Esther Álvarez President of Private Foundation In-nova and Research of the UPM. Consultant in strategic communications programs radio, mobile and wireless both public and private sectors and in civil and military. Currently a member of the board of the Delegation of COIT (Association of Telecommunications Engineers) CLM, representative of Castilla La Mancha in the groups of the free and COIT New Technologies of the National Coordinator of the Treatment Research Chair in Digital Image at the Madrid Polytechnic University of Madrid. PhD in Information Systems specializing in Business ETSI Industriales (UPM) and the Specialty Program Communications Signals, Systems and Radiocommunications Department SSR ETSI Telecomunicaciones (UPM).

David G. Rosado has an MSc and PhD. in Computer Science from the University of Málaga (Spain) and from the University of Castilla-La Mancha (Spain), respectively. His research activities are focused on security for Information Systems and Cloud Computing. He has published several papers in national and international conferences on these subjects, and he is co-editor of a book and chapter books. Author of several manuscripts in national and international journals (Information Software Technology, System Architecture, Network and Computer Applications, etc.). He is member of Program Committee of several conferences and workshops nationals and internationals such as ICEIS, ICCGI, CISIS, SBP, IAS, SDM, SECRIPT, COSE and international journals such as Internet Research, JNCA, KNOSYS, JKUS, and so on. He is a member of the GSyA research group of the Information Systems and Technologies Department at the University of Castilla-La Mancha, in Ciudad Real, Spain.

Eduardo Fernández-Medina holds a PhD. and an MSc. in Computer Science from the University of Sevilla. He is associate Professor at the Escuela Superior de Informática of the University of Castilla-La Mancha at Ciudad Real (Spain), his research activity being in the field of security in databases, datawarehouses, web services and information systems, and also in security metrics. Fernández-Medina is co-editor of several books and chapter books on these subjects, and has several dozens of papers in national and international conferences (DEXA, CAISE, UML, ER, etc.). Author of several manuscripts in national and international journals (Information Software Technology, Computers And Security, Information Systems Security, etc.), he is director of the GSyA research group of the Information Systems and Technologies Department at the University of Castilla-La Mancha, in Ciudad Real, Spain.

SEGURIDAD EN REDES

Análisis y evaluación de la Seguridad proporcionada por DNSSEC en Redes de Información IPv6 en un escenario de pruebas controlado

D. K. Terán Arévalo, D. V. Fernández García, y F. J. Terán Cuarán.

Resumen— El presente artículo se basa en el Análisis y evaluación de la seguridad proporcionada por DNSSEC en redes de información IPv6 en un escenario de pruebas controlado. Para el desarrollo de esta investigación, se implementó un escenario de pruebas controlado para el despliegue de DNSSEC en redes IPv6 en diferentes Ss. Os. (Windows, Linux), con equipos del Departamento de Telecomunicaciones de la Universidad del Cauca (Routers Cisco 2801, Routers software Quagga, Switches Catalyst 2960- S y Pcs). Además, se adaptó una metodología para el análisis de la seguridad proporcionada por DNSSEC en Redes IPv6 en el escenario de pruebas seleccionado, con base en el estudio de las metodologías de evaluación de la seguridad: ISSAF, OSSTMM y PTES, y se evaluó las vulnerabilidades de seguridad de DNSSEC en redes IPv6, con base en el Sistema de puntuación de vulnerabilidad común (CVSS, *Common Vulnerability Scoring System*), mediante diferentes PoC (Pruebas de Concepto), sobre escenarios distintos con diferentes requisitos de seguridad con el fin de determinar en qué casos DNSSEC fue o no fue vulnerable, es decir si DNSSEC proporcionó o no sus funciones de seguridad tales como *autenticación del origen y autenticación e integridad de los datos de DNS*.

Índice de Términos— DNSSEC, DNS, IPv6.

I. INTRODUCCIÓN

El sistema de Nombres de Dominio (DNS, Domain Name System) [1], resuelve un Nombre de Dominio en una dirección IP, por ejemplo el Nombre Dominio *www.google.com* en la dirección IPv6 [2]: *2001:4860:4860::8888*, posibilitando que internet sea más fácil de usar, sin que los usuarios tengan la necesidad de aprender, memorizar y recordar las direcciones IP/IPv6 de cada uno de los Dominios a los que se comunican. Pero aquí, hay un grave problema, todo este proceso es inseguro, ya que en la vida real, un usuario que se conecta desde una terminal y realiza una consulta a un Nombre de Dominio, no sabe quién le envió la respuesta verdadera, y tampoco sabe si los datos fueron modificados por un atacante durante el tránsito, redireccionando a la víctima a un sitio web falso, permitiendo robar de esta manera su información confidencial. Todo esto hace que el DNS sea un protocolo inseguro, debido a que no contempla ningún soporte de seguridad, para garantizar: Autenticación del origen, Autenticación e Integridad de los datos de DNS,

Confidencialidad de la información y Disponibilidad del servicio [3].

Por esta razón la IETF propuso el Sistema de Nombres de Dominio con Extensiones de Seguridad (DNSSEC, Domain Name System Security Extensions) [4, 5, 6], como un conjunto de extensiones que aumentan la seguridad del protocolo DNS, al proveer un mecanismo para la Autenticación del origen e Integridad de los datos DNS, y Denegación de existencia autenticada, mediante el uso de la firma digital de los datos DNS, nuevos registros de recursos y una Infraestructura de Clave Pública (PKI), para la construcción de una “Cadena de Confianza”. Sin embargo, no proporciona la confidencialidad de la información ni disponibilidad del servicio.

En ese sentido, en el presente artículo se presenta el proceso que se llevó a cabo, para Analizar y Evaluar la Seguridad Proporcionada por DNSSEC en Redes de Información IPv6 en un escenario de pruebas controlado, así como los aportes y conclusiones obtenidos como resultado de la investigación.

II. ESCENARIO DE PRUEBAS CONTROLADO

Para determinar el escenario de pruebas controlado, primero se llevó a cabo un estudio detallado sobre el funcionamiento de los Servicios de Internet DNS, DNSSEC y Web en Redes IPv6, y los componentes principales dentro de la infraestructura DNS: Servidores DNS autoritarios, Servidor DNS Caché y Clientes y los componentes DNSSEC:

Servidores autoritarios DNSSEC, Servidor DNS Caché Validador y Cliente Validador, y las funciones que cumplen en el proceso de validación para garantizar la autenticidad e integridad de los datos DNS de una respuesta [7].

Con base en esto se determinaron los componentes principales que formarían parte y cumplirían un rol específico para su implementación en redes IPv6 en el escenario de pruebas controlado.

La topología de red del escenario de pruebas controlado en redes IPv6 que se definió para lograr el propósito del proyecto, está conformada por una red interna y una red externa. La red interna, está constituida por la implementación de la delegación segura DNSSEC para los dominios y subdominios internos de la organización Bancodk, y la red externa, conformada por la implementación de la cadena de

Dalia. K Terán Arévalo, daliateran@unicauca.edu.co, Diana. V. Fernández García, dianvf@unicauca.edu.co y Francisco. J Terán Cuarán, fteran@unicauca.edu.co; Universidad del Cauca, Popayán, Colombia.

confianza DNSSEC, en todos los servidores autoritarios involucrados en los niveles de la jerarquía DNS.

La configuración e implementación de la cadena de confianza DNSSEC se divide en 2 estados: el primero, cuando todos los niveles de la jerarquía DNS se encuentran firmados con DNSSEC, a excepción del dominio networks.com y el segundo cuando la cadena de confianza está rota, es decir que el dominio de nivel superior gTLD.COM, no se encuentra firmado con DNSSEC.

La implementación del escenario de pruebas, se llevó a cabo con equipos del Departamento de Telecomunicaciones de la Universidad del Cauca tales como:

Routers cisco 2801, routers software quagga, switches catalyst 2960-s y pcs y se utilizó el protocolo estándar de enrutamiento BGP para redes IPv6 [8].

Para permitir la conectividad de todos los componentes involucrados en el escenario, definiendo 6 sistemas autónomos, para especificar las redes IPv6 del escenario en formato hexadecimal, como se aprecia en la Figura 1.

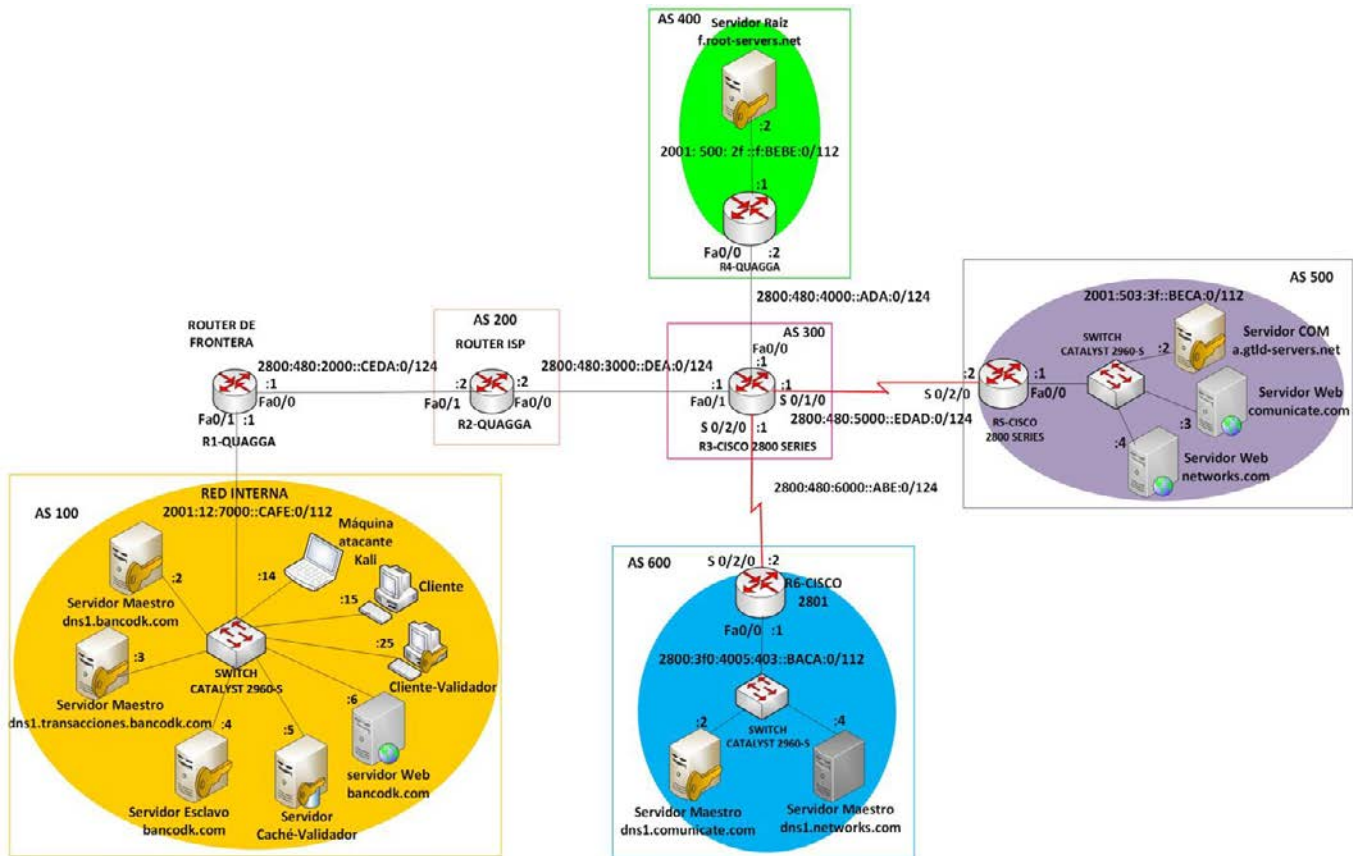


Figura 1. Escenario de pruebas controlado DNSSEC en IPv6.
Fuente: Los autores.

III. METODOLOGÍAS DE EVALUACIÓN DE LA SEGURIDAD

Para adaptar una metodología para el análisis y evaluación de la seguridad proporcionada por DNSSEC en redes IPv6 en el escenario de pruebas controlado, se realizó un estudio, análisis detallado y comparación de los aspectos y características más importantes de las metodologías de evaluación de la seguridad: ISSAF 2.1 [9], OSSTMM 3.0 [10] y PTES 1.1 [11], las cuales tienen en común que el tipo de licencia con el que se han creado permite su utilización de forma libre, es decir sin costo, cuyo resultado fue la definición de la Metodología Adaptada.



Figura 2. Metodologías de evaluación de seguridad.
Fuente: Tomada de [8].

En la Tabla I, se observa la comparación general de los aspectos más relevantes incluidos en cada una de las metodologías, las marcas de verificación de color verde representan los aspectos que tienen en común todas las metodologías, y las marcas de verificación rojas los que no.

TABLA I
COMPARACIÓN GENERAL DE LAS METODOLOGÍAS
DE EVALUACIÓN DE LA SEGURIDAD

ASPECTOS	ISSAF	OSSTMM	PTES
Acreditación			
Alcance			
Introducción a conceptos			
Gestión del compromiso			
Gestión de proyecto			
Entornos de aplicabilidad			
Metodología de penetración			
Guía técnica			
Métricas			
Informes			
Recomendaciones			
Usabilidad y uso			

IV. DEFINICIÓN DE LA METODOLOGÍA ADAPTADA PARA EL ANÁLISIS Y EVALUACIÓN DE LA SEGURIDAD PROPORCIONADA POR DNSSEC

A partir del análisis comparativo de los aspectos más relevantes incluidos en cada una de las metodologías mencionadas anteriormente, se determinaron los aspectos y características más importantes que sirvieron como base para la definición de la Metodología Adaptada, resaltando las características generales que se querían obtener de manera que se adapte a las necesidades del Proyecto para el análisis y evaluación de la seguridad proporcionada por DNSSEC en redes IPv6 en un escenario de pruebas controlado, siendo comprensible, detallada y de fácil aplicabilidad en este contexto.

A. Alcance

El alcance de la Metodología Adaptada se compone de: la definición del test de intrusión, los objetivos, el ambiente o entorno donde se realizarán las pruebas, y la metodología de penetración con su correspondiente guía técnica, métricas y estructura para la entrega de resultados.

B. Entorno de Aplicabilidad

El entorno de aplicabilidad de la Metodología es la seguridad de los sistemas y redes de información, desde la perspectiva de la evaluación de la seguridad proporcionada por la extensión de seguridad del protocolo DNSSEC.

C. Metodología de Penetración

La metodología de penetración de la Metodología Adaptada, está fundamentada en los procedimientos de las metodologías de penetración de PTES e ISSAF.

D. Guía Técnica

La guía técnica de la Metodología Adaptada, se basa en el modelo propuesto por la metodología de penetración de ISSAF, al especificar las herramientas y como se pueden utilizar para la ejecución de las pruebas de cada una de las fases.

E. Métricas

Este trabajo, no cubre el parámetro de métrica como la mayoría de las metodologías estudiadas, mediante el análisis y evaluación del riesgo, sino que se enfoca en el análisis y evaluación de las vulnerabilidades por medio del Sistema Común de Puntuación de Vulnerabilidades (CVSS, Common Vulnerability Scoring System), el cual es un estándar de la industria para evaluar la gravedad de las vulnerabilidades de seguridad informática del sistema.

F. Informes

Los informes de la Metodología Adaptada incluyen los criterios básicos contenidos en los informes de pruebas de penetración de las metodologías de OSSTMM, ISSAF y PTES, con el fin de fortalecer y complementar dichos criterios propuestos.

G. Recomendaciones

Se establecen recomendaciones sobre cómo se deben llevar a cabo cada una de las fases de la Metodología adaptada para analizar y evaluar la seguridad proporcionada por DNSSEC en un escenario de pruebas controlado en redes IPv6, con base en las recomendaciones que proponen las metodologías ISSAF, OSSTMM y PTES.

H. Meticulosidad

El nivel de meticulosidad de la Metodología Adaptada, tiene en cuenta la minuciosidad y exactitud que aportan la metodología ISSAF para el manejo detallado de la guía técnica para la realización de cada una de las pruebas, haciendo uso de herramientas actualizadas como el manejo exhaustivo de OSSTMM para el manejo de la métrica, en nuestro caso orientada a llevar a cabo un correcto análisis de la evaluación de la seguridad teniendo en cuenta el estándar CVSS. De igual forma teniendo en cuenta los criterios contenidos en los reportes y los lineamientos de acción de todas las metodologías estudiadas.

I. Usabilidad y uso

El nivel de usabilidad de la Metodología Adaptada es alto, como resultado a la meticulosidad en la descripción de la planeación, la prueba técnica de penetración, la evaluación de los resultados obtenidos mediante el CVSS y la presentación de informes.

V. FASES DE LA METODOLOGÍA ADAPTADA

Con base en el análisis de las metodologías anteriores, se determinó que la Metodología Adaptada para analizar y

evaluar la seguridad de DNSSEC en Redes de Información IPv6 está conformada por las siguientes cuatro Fases principales, como se observa en la Figura 3:



Figura 3. Fases de la Metodología Adaptada.
Fuente: Los autores.

VI. METODOLOGÍA ADAPTADA PARA ANALIZAR Y EVALUAR LA SEGURIDAD DE DNSSEC EN REDES DE INFORMACIÓN IPV6

A. Fase 1: Planeación

En esta fase, se describen los objetivos, el tipo de test de intrusión, el diagrama de red del ambiente de prueba, el alcance y el plan de pruebas definido que se llevará a cabo en el escenario real de pruebas controlado, como se muestra a continuación:

- 1) **Objetivo:** analizar y evaluar la seguridad proporcionada por DNSSEC en redes de información IPv6 en el escenario real de pruebas controlado.
- 2) **Tipo de test de intrusión:** Caja Gris.
- 3) **Alcance:** Pruebas en las redes interna y externa del escenario con determinados requisitos de seguridad.
- 4) **Diagrama de red:** es el escenario de pruebas controlado, mostrado en la Figura 1.
- 5) **Plan de pruebas:** El plan de pruebas de seguridad se divide en dos tipos de pruebas, el plan de pruebas de recolección de información y el plan pruebas de identificación y explotación de vulnerabilidades, indicando con una X las pruebas que se realizaron en cada escenario, como se muestra en las Tabla II y III.

TABLA II
PLAN DE PRUEBAS DE RECOLECCIÓN DE INFORMACIÓN

	SISTEMA OPERATIVO	EJECUCIÓN DE LA PRUEBA	RED OBJETIVO	PoC	DNS	DNSSEC
RECOLECCIÓN DE INFORMACIÓN	Centos Debian Windows-Debian Windows-Centos	RED INTERNA	INTERNA	Descubrir host activos en segmento de red	X	X
				Encontrar Sistemas Operativos servicios y versiones	X	X
				Identificar relaciones IPv6 con servidores, dominios y subdominios	X	X
			EXTERNA	Identificar relaciones IPv6 con servidores, dominios y subdominios	X	X

TABLA III
PLAN DE PRUEBAS DE EXPLOTACIÓN DE VULNERABILIDADES IDENTIFICADAS

	SISTEMA OPERATIVO	EJECUCIÓN DE LA PRUEBA	SERVIDORES O DOMINIOS INTERNO/ EXTERNO OBJETIVO	POC	DNS	DNSSEC NSEC	DNSSEC NSEC3
IDENTIFICACIÓN Y EXPLOTACIÓN DE VULNERABILIDADES	Centos Debian Windows-Debian Windows-Centos	INTERNA	INTERNA	Transferencia de zona	X	X	X
				Enumeración de dominios por consultas incorrectas		X	X
				Denegación de servicio	X		X
				DNS Spoofing			X
		INTERNA	EXTERNA	Transferencia de zona	X	X	X
				Enumeración de dominios por consultas incorrectas		X	X
				Denegación de servicio	X		X
				DNS Spoofing			X

B. Fase2: Penetración

Esta fase se lleva a cabo las actividades de Recolección de información e Identificación y Explotación de Vulnerabilidades.

- 1) **Identificación de Vulnerabilidades:** Esta actividad se realizó con base en la documentación investigada donde se encontraron una serie de posibles vulnerabilidades del servicio DNS y DNSSEC relacionadas con:
 - El uso de *DNSSEC con RR NSEC*, que permite enumerar el contenido de una zona, a partir de la consulta de Nombres de Dominio que no existen. RFC 5551, RFC 7129 [13], [14].
 - El uso de *UDP como protocolo de transporte*, que posibilita IP spoofing. RFC3833 [12].

- El Resolver DNSSEC no puede verificar las respuestas que se originan en una zona sin firmar. RFC4033 [4].
- Un Cliente resolver que NO realiza validación. RFC4033 [4].

2) **Explotación de Vulnerabilidades:** Para verificar la existencia de vulnerabilidades en DNSSEC, se realizaron las siguientes PoC:

1. Transferencia de zona.
2. Enumeración de dominio por consultas incorrectas.
3. Denegación de Servicio.
4. Secuestro de URL por ataque Typosquatting.
5. DNS Spoofing.

C. Fase 3: Evaluación de vulnerabilidades

La evaluación de las vulnerabilidades de DNSSEC en redes IPv6, se realizó con base en el Sistema de Puntuación de Vulnerabilidad Común CVSS [15], que proporciona una forma de capturar las características principales de una vulnerabilidad y producir una puntuación numérica asociado a un valor de severidad, como se observa en la Tabla IV, y está compuesto por tres grupos principales de métricas: Base, Temporal y de Entorno, cada una con un conjunto de métricas como se observa en la Tabla V:

TABLA IV
PUNTUACIÓN CVSS Y VALOR CUALITATIVO (SEVERIDAD)

Puntuación	Severidad
0	Nula
0.1-3.9	Baja
4.0-6.9	Media
7.0-8.9	Alta
9.0-10.0	Crítica

TABLA V
FASE DE EVALUACIÓN DE VULNERABILIDADES

GRUPO METRICAS BASE	GRUPO METRICAS TEMPORAL	GRUPO METRICAS ENTORNO
Métricas de Explotabilidad • Vector de Ataque • Complejidad de Ataque • Requerimiento de Privilegios • Interacción de Usuarios Métricas de Impacto • Impacto de Confidencialidad • Impacto de Integridad	• Madurez del Código de Explotación • Nivel Remediación • Reporte de Confianza	• Requisitos de Confidencialidad • Requisitos de Integridad • Requisitos de Disponibilidad

Se realizó el cálculo CVSS en cada Prueba de Concepto de las actividades de la Fase 2 de Penetración, determinando los valores de cada uno de los parámetros de las métricas Base, Temporal y Ambiental, hasta obtener el resultado final de la severidad de la vulnerabilidad [16]. A continuación se muestra un Ejemplo de los resultados obtenidos:

1) Evaluación de vulnerabilidad con la PoC DNS Spoofing Dominio Inexistente, Cadena de Confianza Firmada

En este ejemplo se realiza el cálculo de la severidad con la PoC de DNS Spoofing exitosa, cuando la cadena de confianza DNSSEC está Firmada y el MITM se realiza entre el Servidor Caché Validador Windows Server y la Gateway de la red CAFE cuando el Cliente No Validador consulta por el Dominio inexistente **coomunicate.com**, y el escenario está implementado sobre el sistema operativo Windows-Centos/Debian (Red interna-Red Externa).

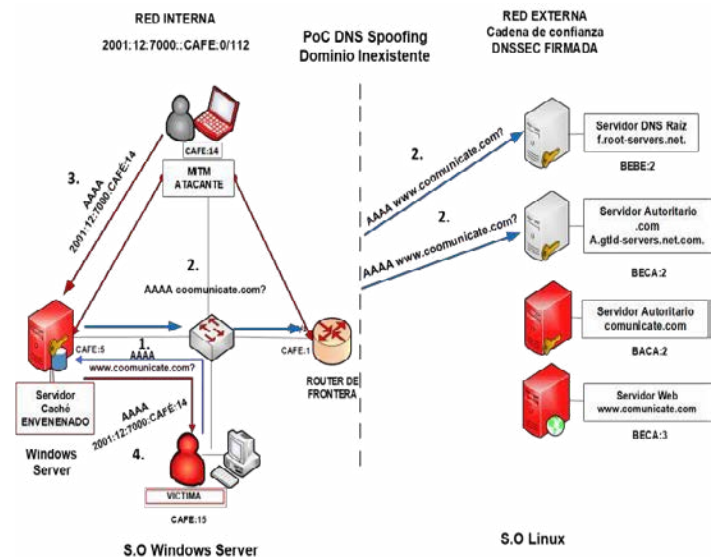


Figura 4. PoC DNS Spoofing, Dominio Inexistente.
Fuente: Los autores.

TABLA VI
PUNTAJE BASE

GRUPO DE MÉTRICAS	MÉTRICA	SIGLA	VALORES POSIBLES	PUNTAJE MÉTRICA
MÉTRICAS BASE	Vector de ataque	AV	Adyacente	7.9
	Complejidad de ataque	AC	Alto	
	Privilegios requeridos	PR	Ninguno	
	Interacción de usuario	UI	Requerido	
	Alcance	S	Cambio	
	Confidencialidad	C	Alto	
	Integridad	I	Alto	
	Disponibilidad	D	Alto	

TABLA VII
PUNTAJE TEMPORAL

GRUPO DE MÉTRICAS	MÉTRICA	SIGLA	VALORES POSIBLES	PUNTAJE MÉTRICA
MÉTRICA TEMPORAL	Explotar el vencimiento del Código	E	Alto	7.7
	Nivel de remediación	RL	solución Alternativa	
	Informe de confianza	RC	Confirmado	

TABLA VIII
PUNTAJE AMBIENTAL

GRUPO DE MÉTRICAS	MÉTRICA	SIGLA	VALORES POSIBLES	PUNTAJE MÉTRICA
MÉTRICAS AMBIENTALES	Requisito de Confidencialidad	CR	Bajo	7.7
	Requisito de Integridad	IR	Alto	
	Requisito de Disponibilidad	DR	Alto	

Una vez asignado los valores de cada uno de los parámetros de las métricas Base, Temporal y Ambiental para esta prueba de concepto, se determina que el valor cuantitativo de la vulnerabilidad está representado por el siguiente puntaje como se puede apreciar en la Figura 5:

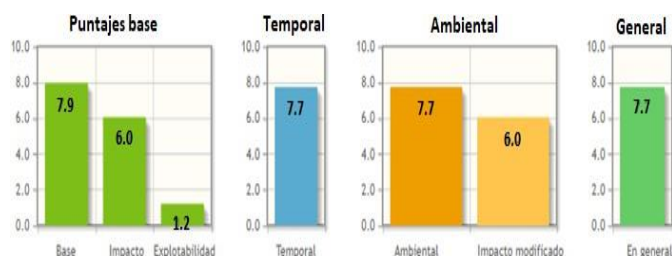


Figura 5. Cálculo Puntaje general CVSS
Fuente: Los autores.

TABLA IX
PUNTUACIÓN CVSS DNS SPOOFING

Puntuación	Severidad
7.7	Alta

Como se obtiene una **Puntuación CVSS** de 7.7 esta corresponde a una **severidad asociada de Media**.

D. Fase 4: Reporte de auditoría

Se divide en 2 reportes:

- Reporte ejecutivo:** tiene como objetivo dar conocer de manera resumida los resultados obtenidos de las PoC, con su respectivo análisis, conclusiones y recomendaciones.
- Reporte técnico:** Se documentan todos los hallazgos de las PoC realizadas sobre el escenario real de pruebas controlado, con una descripción escrita, las herramientas utilizadas, la topología de red, el ambiente real de la prueba, las características de los equipos involucrados por prueba, capturas de pantalla como evidencia y algunas recomendaciones para tener en cuenta.

En las Tablas X, XI y XII, se muestra la clasificación de severidad, con las PoC exitosas en el escenario real de pruebas controlado DNSSEC en redes de información IPv6:

TABLA X
CLASIFICACIÓN DE SEVERIDAD, CON LAS PRUEBAS EXITOSAS DE LA ACTIVIDAD DE RECOLECCIÓN DE INFORMACIÓN

SISTEMA OPERATIVO	EJECUCIÓN DE LA PRUEBA	ESTADO DE LA CADENA DE CONFIANZA DNSSEC	PRUEBA	N° PoC exitosas	SEVERIDAD
Centos/Debian Windows-Debian Windows-Centos	Red Interna	Firmada	Encontrar hosts activos en el segmento de red	1	BAJA
			Encontrar Ss. Os, servicios y versiones	1	MEDIA
			Identificar Relaciones de Direcciones IPv6 con Servidores, Dominios y Subdominios	4	MEDIA
			Total de pruebas		6

TABLA XI
CLASIFICACIÓN DE SEVERIDAD, CON LAS PRUEBAS EXITOSAS DE LA ACTIVIDAD DE IDENTIFICACIÓN Y EXPLOTACIÓN DE VULNERABILIDADES

SISTEMA OPERATIVO	EJECUCIÓN DE LA PRUEBA	ESTADO DE LA CADENA DE CONFIANZA DNSSEC	PRUEBA	N° PoC EXITOSAS	SEVERIDAD
Centos	Red Interna	Firmada	Transferencia de Zona	2	BAJA
	Red Externa			4	BAJA
Centos/Debian Windows-Debian Windows-Centos	Red Interna	Firmada	Enumeración de Dominios por Consultas Incorrectas	2	BAJA
			Denegación de Servicio (DoS)	2	ALTA
			Total de pruebas	10	

TABLA XII
CLASIFICACIÓN DE SEVERIDAD, CON LAS PRUEBAS EXITOSAS DE DNS SPOOFING

SISTEMA OPERATIVO	EJECUCIÓN DE LA PRUEBA	ESTADO DE LA CADENA DE CONFIANZA DNSSEC	ESTADO DEL DOMINIO	PRUEBA	N° PoC EXITOSAS	SEVERIDAD
Debian/Centos	Red Interna	Firmada	F-Int F-ext / NF / NF-I	DNS Spoofing	21	MEDIA
Windows-Centos Windows-Debian			F-Int F-Ext	DNS Spoofing	3	MEDIA
			NF-I	DNS Spoofing	3	ALTA
Centos/Debian Windows-Debian Windows-Centos		Rota	F-Int / NF / NF-I	DNS Spoofing	23	MEDIA
			F-Ext	DNS Spoofing	6	ALTA
Debian	Red Externa	Firmada	F-Ext NF-I	DNS Spoofing	2	ALTA

VII. CONCLUSIONES

- Para analizar y evaluar la seguridad de los Servicios de Internet DNS, DNSSEC y Web en Redes IPv6, se requiere implementar toda la Jerarquía DNS y la Cadena de confianza DNSSEC, en un escenario de pruebas controlado
- La Metodología Adaptada para analizar y evaluar la seguridad de DNSSEC en Redes de Información IPv6 incluye las siguientes las Fases: Planeación, Penetración, Evaluación de Vulnerabilidades y Reporte de Auditoría.

- Implementar DNSSEC en Windows Server, es mucho más complejo que implementar DNSSEC en una Distribución de Linux, se requiere mayor capacidad de recursos y procesamiento, para optimizar el servicio.

- Para la Recolección de Información y Explotación de las Vulnerabilidades relacionadas con la seguridad de DNSSEC en Redes IPv6, actualmente son muy pocas las herramientas existentes y disponibles.

Con base en la **clasificación de severidad**, con las PoC de DNS Spoofing, **se determinó que los peores escenarios en que DNSSEC NO proporciona sus funciones de seguridad:** autenticación del origen, y autenticación e integridad de los datos de DNS.

- Cuando la **cadena de confianza está firmada**, el peor escenario presenta una clasificación de **severidad Media**, cuando el proceso de **validación DNSSEC es realizado únicamente por el Servidor Caché o cuando se consulta por un dominio NO firmado**, debido que bajo este escenario, **el cliente que no realiza el proceso de validación es vulnerable a la suplantación de dominio** al aceptar la respuesta falsa del atacante.
- Cuando la **cadena de confianza está rota**, el peor escenario presenta una clasificación de **severidad Alta**, cuando **se consulta por un dominio firmado y el cliente como el servidor cache que realicen el proceso de validación DNSSEC no tienen almacenada el ancla de confianza del dominio**, de modo que no se puede realizar la validación de la respuesta.

Con base en las **PoC NO exitosas** de DNS Spoofing, **se determinó que los mejores escenarios en que DNSSEC PROPORCIONA sus funciones de seguridad:** autenticación del origen, y autenticación e integridad de los datos de DNS.

- Cuando la **cadena de confianza está firmada**, el mejor escenario se presenta cuando el **Cliente como el Servidor Caché con soporte de validación DNSSEC, tienen almacenada el ancla de confianza del servidor Raíz y/o del dominio interno de la organización**, de esta manera son capaces de validar las respuestas DNS y evitan la suplantación de identidad de un dominio y el envenenamiento de la memoria caché.
- Cuando la **cadena de confianza está rota**, el mejor escenario se presenta cuando el **Cliente como el Servidor Caché con soporte de validación DNSSEC, tienen almacenada el ancla de confianza de un dominio firmado con DNSSEC**.
- DNSSEC NO protege contra ATAQUES de:

1. Denegación de Servicio (DoS, Denial of Service).
2. Transferencia de zona.
3. MITM (Man in the Middle).

REFERENCIAS

- [1] R. Arends, "Internet Protocol, Version 6 (IPv6) Specification", IETF, RFC 2460, Diciembre 1998.
- [2] P. Mockapetris, "Domain Names - Implementation And Specification", IETF, RFC 1035, November 1987.R.
- [3] C. Gónzales, "Análisis de vulnerabilidades del DNS", Proyecto de Grado, Ingeniería de Sistemas Telemáticos, E.T.S.I. Telecomunicación (UPM), Madrid, 2015.
- [4] Arends, "DNS Security Introduction and Requirements", IETF, RFC 4033, Marzo 2005.
- [5] R. Arends, "Resource Records for the DNS Security Extensions", IETF, RFC 4034, Marzo 2005.
- [6] R. Arends, "Protocol Modifications for the DNS Security Extensions", IETF, RFC 4035, Marzo 2005.
- [7] T. E. Sánchez, "Extensiones de Seguridad para el Sistema de Nombres de Dominio", Trabajo de Grado, Facultad de informática, Universidad Nacional de la Plata, UNLP, Argentina, 2012.
- [8] Y. Rekhter, "IPv6 Address Specific BGP Extended Community Attribute", IETF, RFC 5701, Noviembre 2009.
- [9] Rathore and Oissg, "ISSAF-Information Systems Security Assessment Framework 0.2.1B," p. 845, 2006.
- [10] Herzog, Pete, "OSSTMM: The Open Source Security Testing Methodology Manual: v3", 2010.
- [11] Nickerson, D. Kennedy, and C. J. Reil, "The Penetration Testing Execution Standard," 2014, 2017.
- [12] Atkins, "Threat Analysis of the Domain Name System", IETF, RFC 3833, Agosto 2004.
- [13] R. Gellens, Ed, "Lemonade Notifications Architecture", IETF, RFC 5551, Agosto 2009.
- [14] R. Gieben, "Authenticated Denial of Existence in the DNS", IETF, RFC 7129, Febrero 2014.
- [15] T. Base and T. Base, "Common Vulnerability Scoring System v3.0 Examples," no. July, pp. 1–38, 2016.
- [16] FIRST.org. (1989). Common Vulnerability Scoring System Version 3.0 Calculator [Online]. Available: <https://www.first.org/cvss/calculator/3.0> FIRST.org

Dalia Kelly Terán Arévalo Ingeniera en Electrónica y Telecomunicaciones de la Universidad del Cauca. Candidata a Magister en Seguridad Informática de UNIR “Universidad Internacional de la Rioja - España”. Áreas de interés: Seguridad en redes y sistemas de información, protocolos de routing en IPv4 e IPv6, Redes IPv4/IPv6 cableadas e inalámbricas.

Diana Victoria Fernández García Ingeniera en Electrónica y Telecomunicaciones de la Universidad del Cauca. Áreas de interés: lenguajes de programación C, Linux y MatLab. Experiencia en la configuración e instalación de servidores Windows, Linux y Centos.

Francisco Javier Terán Cuarán Magíster en Electrónica y Telecomunicaciones, Universidad del Cauca. Ingeniero en Electrónica, Universidad del Cauca. Pasantía Internacional en Redes y Servicios Telemáticos, Universidad de Chile. Gestor, Promotor y Coordinador del Programa de Especialización en Redes y Servicios Telemáticos de la Universidad del Cauca. Gestor y Promotor del Programa de Ingeniería de Sistemas de la Universidad del Cauca. Gestor, Promotor y Coordinador de la Red de Información de la Universidad del Cauca. Gestor y Promotor del Programa de Maestría en Telecomunicaciones de la Universidad del Cauca. Profesor Titular de Planta adscrito al Departamento de Telecomunicaciones de la FIET de la Universidad del Cauca. Profesor en pregrado y posgrado de la FIET. Decano de la Facultad de Ingeniería Electrónica y Telecomunicaciones (FIET) de la Universidad del Cauca. Jefe del Departamento de Telecomunicaciones de la FIET. Jefe de la División de Tecnologías de la Información y las Comunicaciones (TIC) de la Universidad del Cauca. Auditor interno de Sistemas de Gestión de Seguridad de la Información ISO 27001, Certificación de Bureau Veritas. Áreas de interés: Redes y Servicios IPv4. Enrutamiento IPv4/IPv6, Internetworking, Redes y Servicios IPv6, IPv4/IPv6 móvil, Seguridad en Redes y Sistemas de Información, SGSI Sistemas de Gestión de la Seguridad de la Información y Datacenters.

Vulnerability Analysis Oriented to Business Logic of Web Applications. A Case Study

William-Rogelio Marchand-Niño¹
william.marchand@unas.edu.pe

Ana Claudia Terrones Murga¹
ana.terrones@unas.edu.pe

Resumen— El objetivo de este artículo es mostrar un proceso de análisis de vulnerabilidades de aplicaciones web orientada a la lógica de negocio, donde además de la identificación de vulnerabilidades tecnológicas, se debe prestar atención en aquellas relacionadas a las reglas de negocio que no son detectadas por herramientas automatizadas pero que generalmente son de mayor impacto, por lo que es importante la capacidad de análisis del pentester para identificarlas. Este tipo de vulnerabilidades están contempladas en la Guía de Pruebas OWASP v4 (*Open Web Application Security Project*) pero su aplicación es específica y diferenciada en cada situación particular. Para el caso de estudio de la investigación se consideraron algunas aplicaciones de una universidad con autorización previa.

Palabras clave— *pentesting, aplicaciones web, análisis de vulnerabilidades, cvss, owasp.*

I. INTRODUCCIÓN

En un proceso de análisis de vulnerabilidades a través de ataques informáticos autorizados, busca verificar la existencia de vulnerabilidades a nivel de hardware, software y personas, para una posterior explotación para evaluar su impacto y determinar las acciones correctivas. El resultado ayuda a estimar un plan de mitigación para cerrar estas brechas en previsión de un ataque real, y así evitar pérdidas financieras y de información. [1]. También es importante resaltar que la información sobre las vulnerabilidades encontradas durante la prueba de penetración es confidencial y no debe divulgarse hasta que se resuelva por completo [2]. Abordar un análisis de vulnerabilidades y pruebas de penetración puedan darse en un contexto de caja negra (sin información del objetivo), caja blanca (con información completa del objetivo) o caja gris (con información parcial del objetivo) [3].

Asimismo, las organizaciones en estos tiempos no pueden sentirse seguros en el ciberespacio (Internet), puesto que no están libres de ataques contra sus infraestructuras, aplicaciones, usuarios y diversos sistemas digitales [4], por lo que es importante e imprescindible analizar el nivel de vulnerabilidades que puedan presentar los componentes de los sistemas de información (infraestructura tecnológica, aplicaciones) que soportan y procesan los activos de información, además de valorar el impacto que un ataque causaría en los mencionados sistemas [5].

Es importante señalar que existen bases metodológicas para aplicar un análisis de vulnerabilidades tales como:

- Open Web Application Security Project (OWASP)
- Open Source Security Testing Methodology Manual (OSSTMM)

- Vulnerability Assessment and Mitigation (VAM)

El trabajo de investigación se orienta principalmente al análisis de vulnerabilidades no tecnológicas, es decir las relacionadas a la lógica de negocio, que implica un conocimiento preciso de las funciones de las aplicaciones y sitios web en el sentido de responder a la pregunta ¿qué es lo que hace la aplicación y qué no debería hacer? Para responder a esta interrogante se realizaron pruebas de identificación y comprobación de las vulnerabilidades de forma manual con ayuda de algunas herramientas. En estos casos las herramientas automatizadas no son de gran ayuda para la detección de vulnerabilidades orientadas a la lógica de negocio.

Para llevar a cabo las pruebas se consideró la aplicación de los lineamientos de la Guía OWASP, OSSTMM y VAM.

En las siguientes secciones se describen con mayor detalle las metodologías, el sistema de puntuación CVSS para determinar el nivel de severidad de las vulnerabilidades, el detalle de las pruebas aplicadas, los resultados y las conclusiones.

II. METODOLOGÍAS PARA EL ANALISIS DE VULNERABILIDADES

En esta sección se describen brevemente las metodologías utilizadas en el trabajo de investigación.

A. Open Web Application Security Project (OWASP)

Es un proyecto abierto dedicado a identificar y combatir las causas que hacen que el software sea inseguro, y están plasmados en documentos como la Guía de Pruebas OWASP y el Top 10 de OWASP ampliamente adoptados [6] [7]. El método de prueba para aplicaciones web se basa en dos fases:

- Fase Pasiva: Es una prueba para comprender la lógica de la aplicación y verificar elementos que podrían significar una puerta abierta para un análisis detallado. Las herramientas se pueden utilizar para capturar información, por ejemplo, un proxy HTTP que observa todas las solicitudes y respuestas HTTP. Al final de esta fase, el *pentester* conocerá todos los puntos de acceso de la aplicación, tales como encabezados HTTP, parámetros, cookies, etc.
- Fase Activa: En esta fase, el *pentester* comienza a realizar las pruebas utilizando la metodología descrita en la Guía. El conjunto de pruebas activas se ha dividido en 11 categorías para un total de 91 controles.

¹Grupo de Investigación en Redes, Seguridad y Gestión de TI
Universidad Nacional Agraria de la Selva - Tingo María, Perú

B. Open Source Security Testing Methodology Manual (OSSTMM)

El Manual de Metodología de Pruebas de Seguridad de Código Abierto (OSSTMM) es una metodología para probar la seguridad operativa de ubicaciones físicas, flujo de trabajo, pruebas de seguridad humana, pruebas de seguridad física, pruebas de seguridad inalámbricas, pruebas de seguridad de telecomunicaciones, pruebas de seguridad de redes de datos y cumplimiento [8].

El análisis de vulnerabilidades en aplicaciones web se pueden realizar de forma automatizada o manual con apoyo de algunas herramientas. Entre las herramientas más utilizadas para ejecutar pruebas automatizadas en el contexto de un escaneo y análisis de vulnerabilidades se incluyen a OWASP ZAP Proxy, Brup Suite, Acunetix, Nessus, entre otras [9] [10] [11].

La identificación de vulnerabilidades tecnológicas con el uso de herramientas automatizadas no es suficiente, por lo que se debe recurrir a un análisis manual que implica generalmente el conocimiento de la lógica de negocio en las aplicaciones web y que muchas veces pueden tener un mayor impacto y su solución puede ser más compleja [6] [9].

Pero ¿qué es una vulnerabilidad? Una vulnerabilidad es una falla en un sistema, que permite a un atacante comprometer su confidencialidad, integridad o disponibilidad. Las pruebas de penetración de aplicaciones web implican técnicas que conducen a la identificación de vulnerabilidades potenciales, que pueden comprometer las aplicaciones web [9]. Las vulnerabilidades de aplicaciones pueden catalogarse de acuerdo con la Guía OWASP, siendo las más comunes las siguientes [6] [12]:

- SQL Injection (SQLi)
- Cross site scripting (XSS), existen dos tipos básicos: XSS reflejado y XSS almacenado.
- Escalamiento de Privilegios, tanto horizontal como vertical.
- Secuestro de sesiones, generalmente por medio cookies.
- Insuficiente tiempo de expiración de sesiones.
- Directorios transversales. Acceso no autorizado a directorios y archivos no autorizados.
- Fallas en la validación de entradas.
- Referencia Directa Insegura a Objetos (IDOR)
- Clickjacking. Técnica maliciosa para engañar a un usuario para que haga clic en páginas web aparentemente inofensivas.

C. Vulnerability Assessment and Mitigation (VAM)

Para el análisis de vulnerabilidades también existe otra metodología como VAM (*Vulnerability Assessment and Mitigation*) que establece seis etapas generales [13]:

- Etapa 1. Identificar las funciones de información esenciales de la organización.
- Etapa 2. Identificar los sistemas de información esenciales para implementar las funciones esenciales identificadas en la etapa 1.

- Etapa 3. Identificar las vulnerabilidades de los sistemas esenciales identificados en la etapa 2.
- Etapa 4. Identificar las técnicas de seguridad pertinentes para mitigar las vulnerabilidades identificadas en la etapa 3.
- Etapa 5. Seleccionar y aplicar las técnicas de la etapa 4 basadas en restricciones, costos y beneficios.
- Etapa 6. Probar las técnicas aplicadas en la etapa 5 para determinar la robustez y las posibilidades reales bajo amenaza.

De las seis etapas planteadas, las que se alinean con el procedimiento empleado en el trabajo de investigación fueron las etapas 3, 4, 5 y 6, puesto que el objetivo del estudio fue evidenciar las vulnerabilidades de tipo no tecnológico, más allá de realizar un análisis de los activos y sistemas de información que bien puede ser parte de los trabajos futuros.

III. MEDICIÓN DE SEVERIDAD DE VULNERABILIDADES CON CVSS

Para la medición de la severidad de las vulnerabilidades se hace uso del CVSS (Common Vulnerability Scoring System) que es un framework que incluye las características más significativas para evaluar las vulnerabilidades de software y hardware. Este framework ha sido desarrollado por First.org Inc. cuya finalidad es la de ayudar a los equipos de respuesta a incidentes de seguridad informática.

El CVSS consta de tres grupos de métricas, las cuales son, métricas base, métricas temporales y métricas ambientales. El grupo Base representa las cualidades intrínsecas de una vulnerabilidad que son constantes en el tiempo, el grupo Temporal refleja las características de una vulnerabilidad que cambian con el tiempo y el grupo Ambiental representa las características de una vulnerabilidad que son exclusivas en el entorno de una organización [14].

Para esta investigación, la determinación de la puntuación de severidad se aplicó el grupo de métricas base que se componen de las métricas de explotabilidad y las métricas de impacto (ver Fig. 1). El resultado final produce una puntuación que oscila entre 0 y 10 y la categorización se puede observar en la Tabla I.

TABLA I. NIVELES DE SEVERIDAD CVSS

Nivel de severidad de vulnerabilidad	Puntuación CVSS
Ninguno	0.0
Bajo	0.1 – 3.9
Medio	4.0 – 6.9
Alto	7.0 – 8.9
Crítico	9.0 – 10.0

El componente de las métricas de explotabilidad refleja la facilidad de explotación de una vulnerabilidad y el componente de métricas de impacto expresan la consecuencia directa de la explotación exitosa de la vulnerabilidad.

A continuación, se explican brevemente las métricas utilizadas en la determinación del nivel de severidad de las vulnerabilidades identificadas.

- **Vector de ataque.** Esta métrica refleja el contexto por la que es posible la explotación de una vulnerabilidad (red, adyacente, local o físico). El valor será mayor

cuanto más remoto sea posible explotar la vulnerabilidad.

- **Complejidad de ataque.** Describe las condiciones que deben existir para explotar la vulnerabilidad. Puede ser baja si no existen condiciones de acceso especializadas, o alta si requiere que el atacante invierta una cantidad considerable de esfuerzo en la preparación y ejecución contra el componente vulnerable.
- **Privilegios requeridos.** Refleja el nivel de privilegios que debe tener un atacante antes de explotar con éxito la vulnerabilidad. Las alternativas para esta métrica son ninguno, bajo o alto.
- **Interacción con el usuario.** Esta métrica determina si la vulnerabilidad puede explotarse exitosamente a voluntad del atacante o requiere la interacción de un usuario de alguna manera como descargar algún archivo, hacer click en algún enlace, ejecutar o abrir algún documento, etc.
- **Alcance.** Esta métrica determina si una vulnerabilidad explotada impacta o no en recursos que están más allá del componente vulnerable y el alcance de su seguridad.
- **Confidencialidad.** Esta métrica mide el impacto en la confidencialidad de los activos de información debido a una vulnerabilidad explotada con éxito. El posible valor es alto, bajo o ninguno.
- **Integridad.** Esta métrica mide el impacto en la integridad de los activos de información producto de la explotación exitosa de una vulnerabilidad. El valor que puede tomar es alto, bajo o ninguno.
- **Disponibilidad.** Esta métrica mide el impacto en la disponibilidad de los activos de información producto de la explotación exitosa de una vulnerabilidad. El valor que puede tomar es alto, bajo o ninguno.

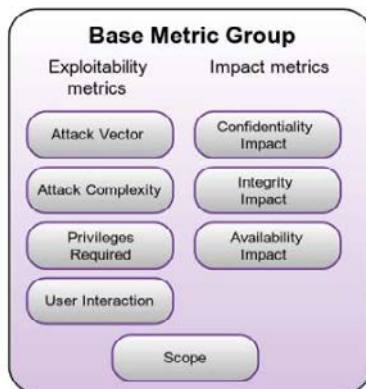


Fig. 1. Grupo de métricas base [14]

IV. DESARROLLO DEL ANÁLISIS DE VULNERABILIDADES ORIENTADO A LA LÓGICA DE NEGOCIO.

Las pruebas presentadas a continuación fueron realizadas según la guía de pruebas OWASP v4 y están orientadas a evaluar la lógica de negocio en las aplicaciones y sitios web.

La Tabla II muestra la lista de controles que se evaluaron en las aplicaciones web.

TABLA II. PRUEBAS REALIZADAS EN LAS APLICACIONES WEB

ID de Prueba	Nombre de la Prueba
Recopilación de información	
OTG-INFO-005	Revisión de los comentarios de la página web y los metadatos para detectar fugas de información.
Pruebas de gestión de identidad	
OTG-IDENT-004	Prueba de enumeración de cuenta de usuario predecible
Pruebas de autenticación	
OTG-AUTHN-001	Prueba de credenciales transportadas a través de un canal encriptado
OTG-AUTHN-002	Prueba de credenciales por defecto
OTG-AUTHN-003	Prueba de mecanismo de bloqueo débil
OTG-AUTHN-004	Pruebas para eludir el esquema de autenticación
Pruebas de gestión de sesión	
OTG-SESS-002	Prueba de atributos de cookies
OTG-SESS-003	Pruebas para la fijación de sesión
OTG-SESS-006	Prueba de funcionalidad de cierre de sesión
OTG-SESS-007	Tiempo de espera de sesión de prueba

TABLA III. PRUEBAS REALIZADAS EN LOS SITIOS WEB

ID de Prueba	Nombre de la Prueba
Configuración e implementación de pruebas de gestión	
OTG-CONFIG-005	Enumerar interfaces de administración de aplicaciones e infraestructura
Pruebas de autenticación	
OTG-AUTHN-003	Prueba de mecanismo de bloqueo débil

Realizadas las pruebas de acuerdo con la Tabla II y Tabla III, se identificaron como vulnerabilidades orientadas a la lógica de negocio a los siguientes hallazgos:

A. Credenciales débiles (OTG-AUTHN-002)

En muchas situaciones, cuando se crea una cuenta en una aplicación, se genera una contraseña predeterminada con algunas características estándar. Si esta contraseña es predecible y el usuario no la cambia en el primer acceso, esto puede hacer que un atacante obtenga acceso no autorizado a la aplicación. Este es el caso de una aplicación denominada Gestión Docente, cuya cuenta de administrador posee una contraseña débil (docente/password) tal como se muestra en la Fig. 2.

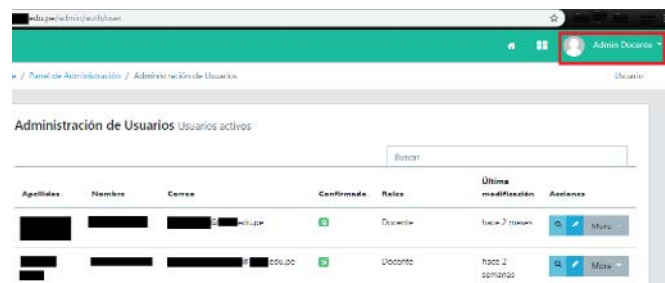


Fig. 2. Sesión con privilegios de administrador en la aplicación de Gestión Docente.

La puntuación determinada en el sistema CVSS es de 7.5 mostrada en la Tabla IV, lo cual indica un nivel de severidad alto, los factores (métricas) que contribuyen a una explotación exitosa de esta vulnerabilidad son: El vector de ataque es la red, es decir, es explotable de manera remota. No existen condiciones especializadas para su explotación, solo fue suficiente realizar una adecuada fase de reconocimiento donde se recolectaron correos electrónicos que sirvieron para elaborar un diccionario y encontrar alguna contraseña débil o por defecto. No se requirió ningún tipo de acceso ni privilegio

antes del ataque. Se puede realizar la explotación sin requerir la interacción de algún usuario. Hay una pérdida total de la confidencialidad, ya que con la explotación se logró el acceso como administrador quedando expuesta información confidencial de todos los usuarios de la aplicación.

TABLA IV. SEVERIDAD DE LA VULNERABILIDAD CREDENCIALES DÉBILES

Métrica	Valor
Vector de ataque	Red
Complejidad de ataque	Baja
Privilegios requeridos	Ninguno
Interacción con el usuario	Ninguno
Alcance	Sin cambios
Confidencialidad	Alto: Pérdida total de confidencialidad
Integridad	Ninguno
Disponibilidad	Ninguno
Puntuación Base CVSS v3.0	7.5

Para mitigar esta vulnerabilidad de contraseñas fáciles de predecir que facilitan el acceso no autorizado, existen dos soluciones: introducir controles de autenticación adicionales (es decir, autenticación de dos factores) o introducir una política de contraseña segura. El más simple y barato de estos es la introducción de una política de contraseñas sólida que garantice la longitud, complejidad, reutilización y antigüedad de las contraseñas, sin embargo, lo recomendable es agregar un factor adicional de autenticación (token, mensaje de texto, llamada por teléfono)

B. Mecanismo de bloqueo débil (OTG-AUTHN-003)

En esta prueba se verificó que el mecanismo de bloqueo no es efectivo, ya que presenta un temporizador aleatorio el cual no impide seguir realizando intentos, lo cual nos indica que la aplicación es susceptible a ataques de diccionario y fuerza bruta.

En la aplicación evaluada, las cuentas se bloquean después de 5 intentos fallidos de inicio de sesión y solo se pueden desbloquear después de un periodo de tiempo predeterminado o por intervención de un administrador. En los resultados de la prueba, que se muestran a continuación, se puede observar que en el sexto intento de inicio de sesión con un usuario válido y un password incorrecto, la aplicación envía un mensaje de bloqueo de la cuenta por 39 segundos (ver Fig. 3), luego, sin esperar el tiempo indicado se realizó un nuevo intento (séptimo intento) con los mismos datos y ahora se muestra un mensaje pero con un tiempo menor al anterior (Fig. 4), sucediendo lo mismo en el octavo intento (Fig. 5) y en el noveno intento ya no se mostró ningún mensaje de bloqueo (Fig. 6).

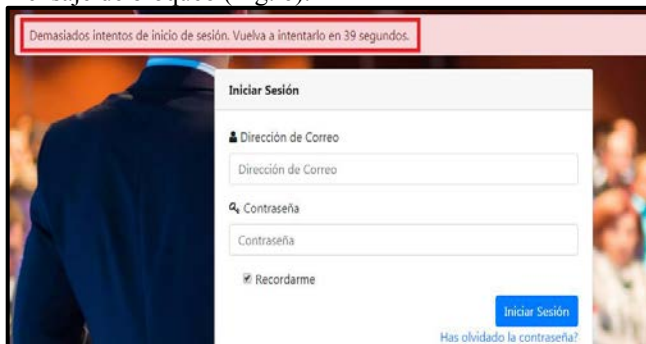


Fig. 3. Sexto intento fallido de inicio de sesión con usuario válido y contraseña incorrecta

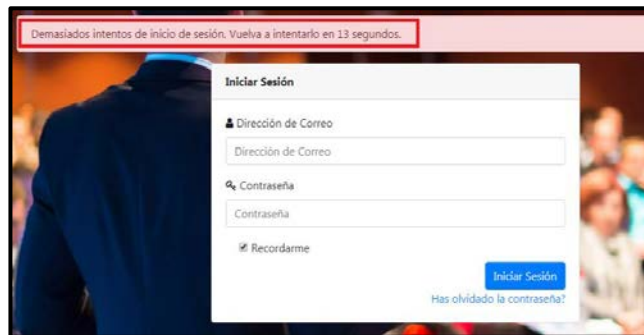


Fig. 4. Séptimo intento fallido de inicio de sesión con usuario válido y contraseña incorrecta

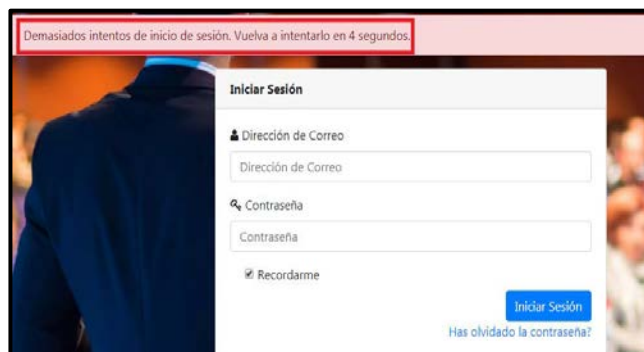


Fig. 5. Octavo intento fallido de inicio de sesión con usuario válido y contraseña incorrecta

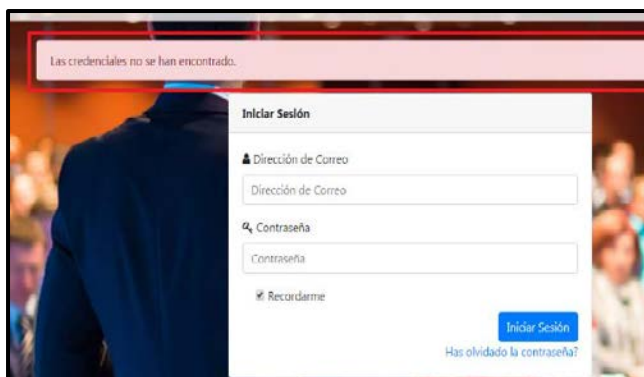


Fig. 6. Noveno intento fallido de inicio de sesión con usuario válido y contraseña incorrecta

La puntuación determinada en el sistema CVSS es de 7.5 tal como se muestra en la Tabla V, lo cual indica un nivel de severidad alto. Los factores (métricas) que contribuyen a una explotación exitosa de esta vulnerabilidad son, el vector de ataque es la red, que significa que es posible su explotación de forma remota. La complejidad de ataque es baja, porque no se requieren condiciones de acceso especiales para su explotación, solo realizar un ataque de *password spraying* para evadir el mecanismo de bloqueo. No se requiere ningún tipo de acceso ni privilegio antes del ataque. Se puede realizar la explotación sin requerir la interacción de algún usuario. En caso de una explotación exitosa, es decir, encontrar credenciales de acceso válidas, significaría una pérdida total de la confidencialidad, porque el atacante tendría acceso a información confidencial de los usuarios de la aplicación.

TABLA V. SEVERIDAD DE LA VULNERABILIDAD MECANISMO DE BLOQUEO DÉBIL

Métrica	Valor
Vector de ataque	Red
Complejidad de ataque	Baja
Privilegios requeridos	Ninguno
Interacción con el usuario	Ninguno
Alcance	Sin cambios
Confidencialidad	Alto
Integridad	Ninguno
Disponibilidad	Ninguno
Puntuación Base CVSS v3.0	7.5

Para mitigar esta vulnerabilidad se recomienda aplicar mecanismos de desbloqueo de cuentas en función del nivel de riesgo. En orden de menor a mayor seguridad: bloqueo y desbloqueo basado en el tiempo; desbloqueo de autoservicio (envía un correo electrónico de desbloqueo a la dirección de correo electrónico registrado); o desbloqueo manual por parte del administrador.

C. Enumeración de usuarios y cuenta de usuario predecible (OTG-IDENT-004)

A menudo, las aplicaciones web revelan cuándo existe o no un nombre de usuario en el sistema, ya sea como consecuencia de una configuración incorrecta o como una decisión de diseño. El propósito de esta prueba es verificar si es posible recopilar un conjunto de nombres de usuario válidos interactuando con el mecanismo de autenticación de la aplicación y los mensajes de errores o advertencias que emiten durante el proceso, o como en este caso de prueba a través del mecanismo de recuperación de contraseña y sus correspondientes mensajes. Un atacante puede utilizar los mensajes que se emiten por el proceso de recuperación de contraseña para obtener una lista de usuarios del sistema que luego pueden formar parte de un diccionario para ataques de fuerza bruta.

En la Fig. 7 se muestra el formulario de recuperación de contraseña donde se solicita la dirección de correo electrónico.

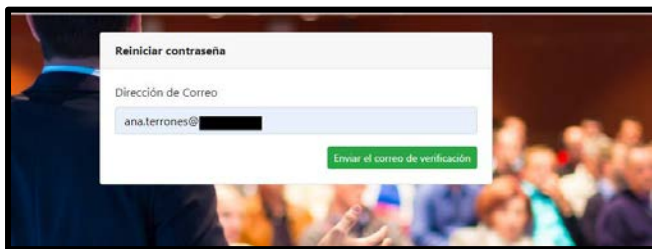


Fig. 7. Formulario de recuperación de contraseña.

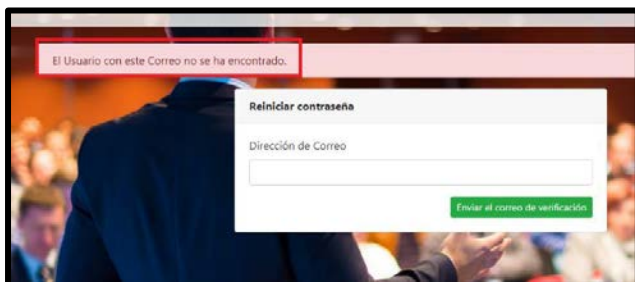


Fig. 8. Mensaje con ingreso de correo no válido.

Al ingresar un correo no válido en el formulario de recuperación de contraseña se muestra el mensaje “El usuario con este Correo no se ha encontrado” tal como se aprecia en la Fig. 8, esto es un indicador de la posibilidad de saber qué cuentas de correo electrónico son válidas o no, y para confirmar que se puede utilizar dichos mensajes para enumerar correos electrónicos válidos, se ingresa un correo válido tal como se muestra en la Fig. 9, y como se puede observar el mensaje “Le hemos enviado el enlace para el reinicio de la contraseña” es distinto al de un correo no válido, lo cual hace posible probar de forma manual o automatizada si una dirección de correo electrónico es válida o no. El propósito final de esta enumeración es la construcción de listas de usuarios válidos para ataques de ingeniería social o



Fig. 9. Mensaje de recuperación de contraseña con ingreso de correo válido.

La puntuación determinada en el sistema CVSS es de 5.3 tal como se muestra en la Tabla VI, lo cual indica un nivel de severidad medio. Los factores (métricas) que contribuyen a una explotación exitosa de esta vulnerabilidad son el vector de ataque (red), es decir, es explotable de manera remota. No existen condiciones especializadas para su explotación, solo fue suficiente realizar una adecuada fase de reconocimiento donde se recolectaron correos electrónicos de otros sitios web del mismo dominio, los cuales sirvieron para enumerar usuarios válidos de la aplicación. No se requiere ningún tipo de acceso ni privilegio para realizar el ataque. Se puede realizar la explotación sin requerir la interacción de algún usuario. La pérdida de confidencialidad es considerada baja, porque el hecho de obtener usuarios del sistema no causa una pérdida directa o grave para la aplicación, pero cabe mencionar que esta vulnerabilidad es un factor agravante para las vulnerabilidades de mecanismo de bloqueo y credenciales débiles, ya que en conjunto ayudan al atacante a crear el escenario ideal para ataques de diccionario o fuerza bruta.

TABLA VI. SEVERIDAD DE LA VULNERABILIDAD DE ENUMERACIÓN DE USUARIOS

Métrica	Valor
Vector de ataque	Red
Complejidad de ataque	Baja
Privilegios requeridos	Ninguno
Interacción con el usuario	Ninguno
Alcance	Sin cambios
Confidencialidad	Baja
Integridad	Ninguno
Disponibilidad	Ninguno
Puntuación Base CVSS v3.0	5.3

Fig. 10. Cookie de la sesión original

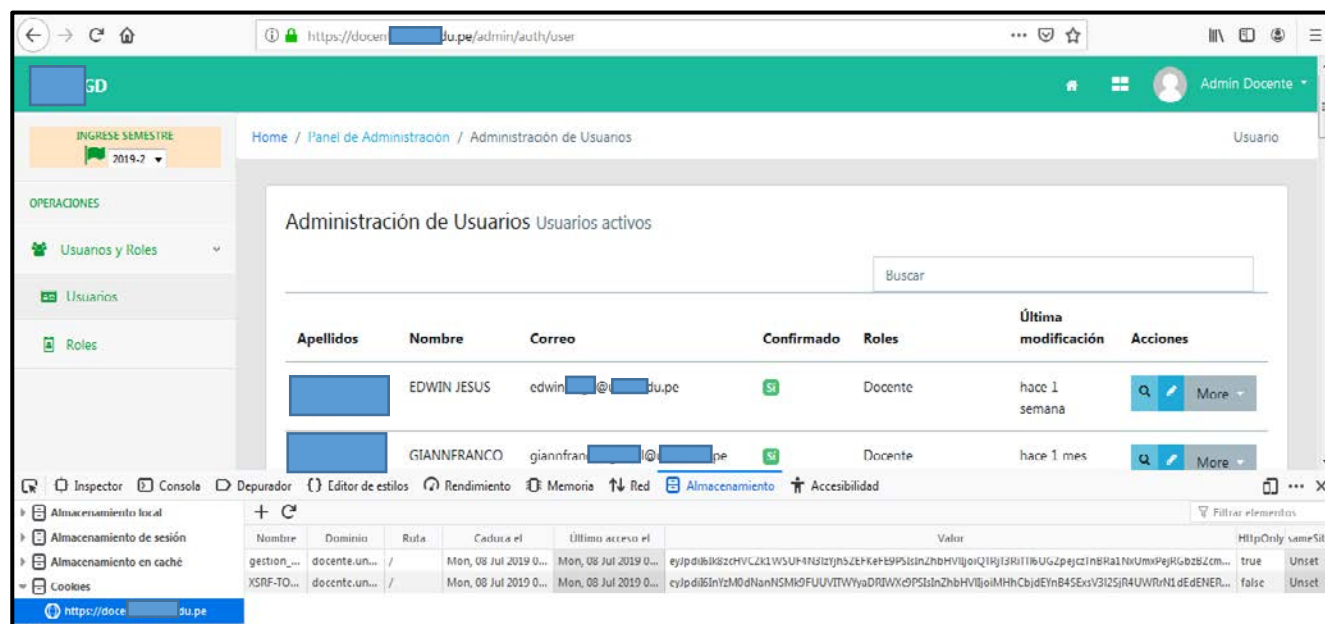
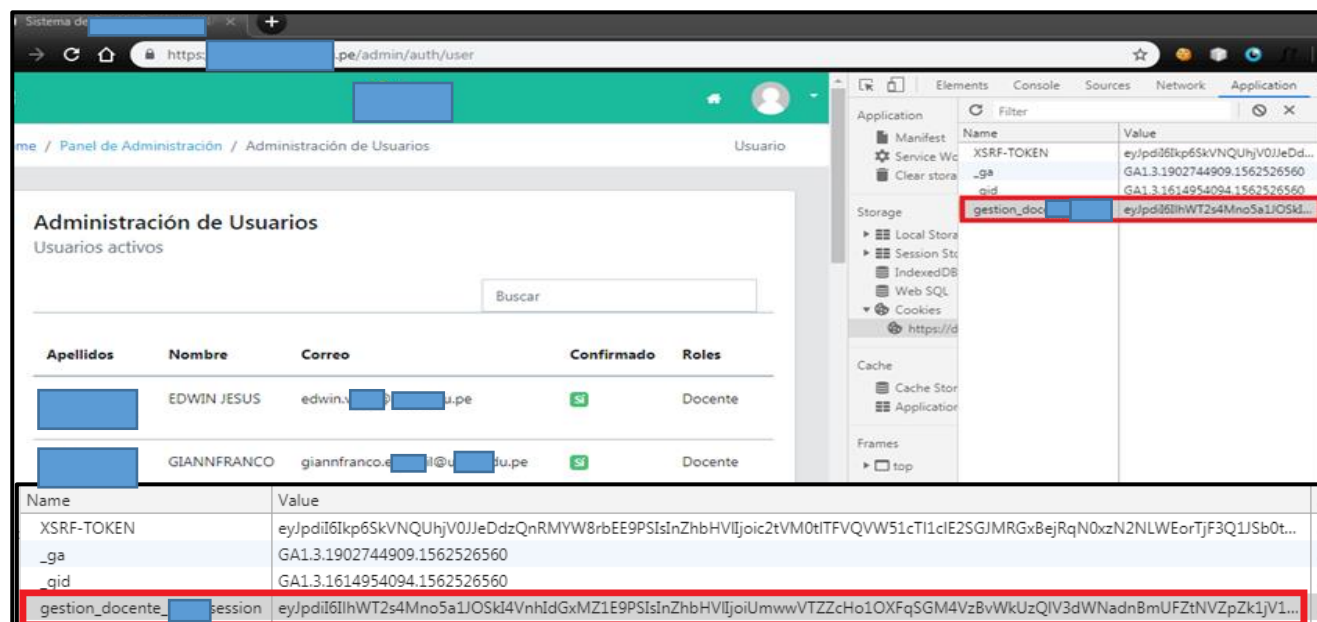


Fig. 11. Reproducción de la sesión reutilización la cookie de sesión en el Sistema de Gestión Docente

D. Reproducción de cookies de autenticación (OTG-SESS-002)

Algunos sitios vulnerables simplemente están utilizando la cookie como token de autenticación sin tener que realizar más comprobaciones en el lado del servidor web, por lo tanto, para comprobar se inició sesión con credenciales válidas, la cual se copió la cookie, en otro navegador, se pega el valor copiado anteriormente en el campo correspondiente en la consola y como resultado se comprueba que es posible iniciar sesión solo con el valor de dicha cookie sin necesidad de ingresar un usuario y una contraseña, lo que hace que el token de autenticación sea vulnerable a los atacantes si las cookies son robadas por *malware* o inyección de JavaScript. Las Fig.

10 y 11 muestran la evidencia en diferentes navegadores.

TABLA VII. SEVERIDAD DE LA VULNERABILIDAD EN LA COOKIE DE SESIÓN

Métrica	Valor
Vector de ataque	Red
Complejidad de ataque	Baja
Privilegios requeridos	Ninguno
Interacción con el usuario	Requerido
Alcance	Sin cambios
Confidencialidad	Bajo
Integridad	Ninguno
Disponibilidad	Ninguno
Puntuación Base CVSS v3.0	4.3

La puntuación en el sistema CVSS es de 4.3, tal como se muestra en la Tabla VII, lo cual indica un nivel de severidad medio, ya que para lograr una explotación exitosa es necesario recrear un escenario donde el atacante capture la cookie y luego logre secuestrar la sesión del usuario.

Se recomienda que además de usar la cookie como token de autenticación, realizar más comprobaciones en el lado del servidor web (validar usuario y contraseña).

E. Funcionalidad de cierre de sesión (OTG-SESS-006)

La terminación de la sesión es una parte importante del ciclo de vida de la sesión. Algunas aplicaciones web se basan únicamente en la cookie de sesión para identificar al usuario que ha iniciado sesión. Al cerrar sesión, la cookie de sesión se elimina del navegador, sin embargo, dado que la aplicación no realiza ningún seguimiento, no sabe si una sesión está desconectada o no. Por lo tanto, al reutilizar una cookie de sesión es posible obtener acceso a la sesión autenticada. Para esta prueba se inició sesión con usuario y contraseña válidos, luego se invocó la función de cierre de sesión y a continuación se navegó a una página que solo está visible en una sesión autenticada, mediante el uso del botón atrás del navegador, volviendo a acceder al panel de administración, lo ideal es que la aplicación redirija al formulario de autenticación, lo cual debería ser necesario para la seguridad de la aplicación, pero el establecimiento de cookies de sesión a nuevos valores después del cierre de sesión generalmente se considera una buena práctica.

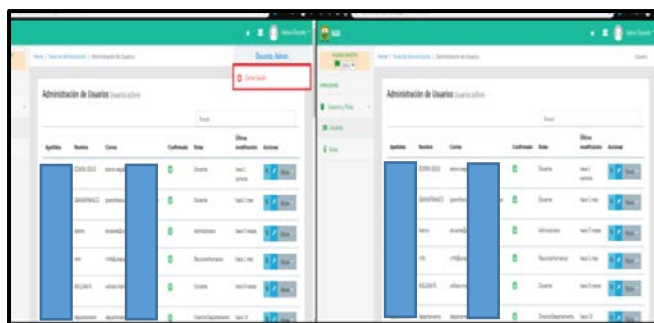


Fig. 12. Prueba de terminación de sesión del lado del servidor en el Sistema de Gestión Docente

Producto de la prueba anterior (reproducción de cookies de sesión) en la que se comprobó que es posible reproducir una sesión usando el mismo valor de la cookie, se verificó la funcionalidad de cierre de sesión cerrando la sesión original (víctima) y se verificó que aún es posible mantener la otra sesión activa, lo cual no implica un alto riesgo, a menos que

un atacante logre capturar la cookie y así logre secuestrar la sesión del usuario, tal como se evidencia en las Fig. 12, Fig. 13 y Fig. 14.

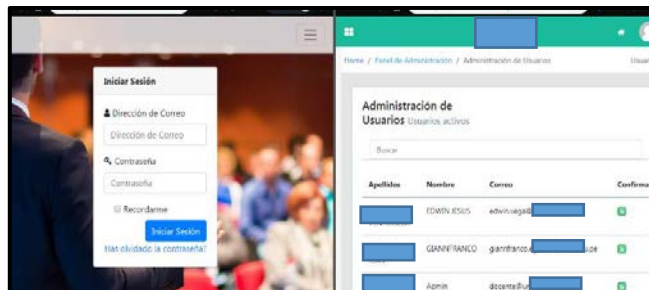


Fig. 13. Prueba de funcionalidad de cierre de sesión

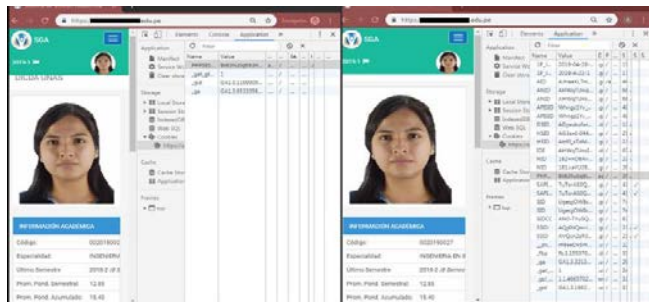


Fig. 14. Reproducción de la sesión utilizando la cookie en el Sistema de Gestión Académica

La puntuación en el sistema CVSS es de 4.3, tal como se muestra en la Tabla VIII, lo cual indica un nivel de severidad medio, ya que para lograr una explotación exitosa es necesario recrear un escenario donde el atacante capture la cookie y luego logre secuestrar la sesión del usuario.

TABLA VIII. SEVERIDAD DE LA VULNERABILIDAD DE LA FUNCIONALIDAD DE CIERRE DE SESIÓN

Métrica	Valor
Vector de ataque	Red
Complejidad de ataque	Baja
Privilegios requeridos	Ninguno
Interacción con el usuario	Requerido
Alcance	Sin cambios
Confidencialidad	Bajo
Integridad	Ninguno
Disponibilidad	Ninguno
Puntuación Base CVSS v3.0	4.3

Se recomienda, que como buena práctica se establezca las cookies de sesión a nuevos valores después del cierre de sesión, además de destruir dicha variable al invocar la función de cierre de sesión. Asimismo, implementar el indicador HttpOnly, que permite enviar cookies solo a través del protocolo http, pero no de JavaScript. Establecer la marca HttpOnly considerando como base el siguiente ejemplo:

```
<?php
ini_set("session.cookie_httponly", "True");
session_start();
session_regenerate_id();
if(!isset($_SESSION['admin_loggedin']))
{
    header('Location: index.php');
}
if(isset($_GET['search']))
{
    if(!empty($_GET['search']))
```

```

{
    $text = $_GET['search'];
}
else
{
    $text = "No text Entered";
}
}
?>

```

Implementar el indicador “secure”, en el encabezado set-cookie para enviar cookies solo a través de HTTPS. Establecer la marca “secure” considerando como base el siguiente ejemplo:

```

<?php
ini_set("session.cookie_secure", "True");
session_start();
session_regenerate_id();
if(!isset($_SESSION['admin_loggedin']))
{
    header('Location: index.php');
}
if(isset($_GET['search']))
{
    if(!empty($_GET['search']))
    {
        $text = $_GET['search'];
    }
    else
    {
        $text = "No text Entered";
    }
}
?>

```

V. RESULTADOS

El análisis de vulnerabilidades presentado como un caso de estudio enfatiza la necesidad de realizar pruebas mas allá de la aplicación de herramientas automatizadas; debido a que como se evidencia, no existe herramienta que detecte y valide la existencia de vulnerabilidades que no sean tecnológicas completamente.

Se ha mostrado que puede existir un impacto significativo con el tipo de vulnerabilidades orientado a la lógica de negocio ya que corresponden a, cómo las aplicaciones funcionan, qué datos transmiten, cómo las transmiten, que privilegios tiene cada rol de una aplicación, manejo de sesiones, manejo de credenciales, etc. Este tipo de fallas están relacionadas a la manera en que se han construido las aplicaciones y sitios web, por lo que la solución por lo general suele ser complicada en términos de tiempo y adaptación.

Las vulnerabilidades identificadas en este caso de estudio tienen un nivel de severidad entre 4.3 (medio) a 7.5 (alto) según CVSS lo cual implica tomar las acciones necesarias para mitigarlas. Comúnmente este tipo de vulnerabilidades están asociadas al manejo de credenciales y sesiones de las aplicaciones.

VI. CONCLUSIONES

En un proceso de análisis de vulnerabilidades se deben considerar tanto las relacionadas a vulnerabilidades tecnológicas como las vulnerabilidades de lógica de negocio, por lo que es importante la capacidad de análisis del pentester o auditor además del manejo de las herramientas adecuadas.

El conocimiento del funcionamiento de las aplicaciones es vital para un análisis exhaustivo y preciso que permita detectar vulnerabilidades importantes que pueden afectar significativamente a la organización. Estas vulnerabilidades pueden ser las más complicadas de identificar.

REFERENCIAS

- [1] B. Babak D. y Z. A. S. Hessa Mohammed, «A study on penetration testing process and tools,» de *Long Island Systems, Applications and Technology Conference (LISAT)*, Farmingdale, NY, 2018.
- [2] Y. Stefinko, A. Piskozub y R. Banakh, «Manual and automated penetration testing. Benefits and drawbacks. Modern tendency,» de *13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET)*, Lviv, Ukraine, 2016.
- [3] R. E. L. d. Jiménez, «Pentesting on web applications using ethical - hacking,» de *2016 IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI)*, San José, 2016.
- [4] S. Smith, «Recognizing and Preparing Loss Estimates from Cyber-Attacks,» *Information System Security*, vol. 12, n° 6, pp. 46-57, 2004.
- [5] P. Aguilera Lopez, Seguridad Informática. Ciclos formativos, Editex, 2010, p. 240.
- [6] Open Web Application Security Project, «OWASP Testing Guide v4,» Abril 2016. [En línea]. Available: <https://www.owasp.org/images/1/19/OTGv4.pdf>.
- [7] Open Web Application Security Project, «OWASP Top 10 2017,» 20 noviembre 2017. [En línea]. Available: <https://www.owasp.org/images/5/5e/OWASP-Top-10-2017-es.pdf>.
- [8] Open Web Application Security Project, «Penetration testing methodologies,» 19 Febrero 2019. [En línea].
- [9] N. Sangeeta y S. Kurkure, «Vulnerability Assessment and Penetration Testing of Web Application,» de *2017 International Conference on Computing, Communication, Control and Automation (ICCCUBE)*, Pune, 2017.
- [10] Open Web Application Security Project, «OWASP ZAP 2.8,» 20 Marzo 2019. [En línea]. Available: https://www.owasp.org/index.php/OWASP_Zed_Attack_Proxy_Project.
- [11] SANS Institute Reading Room site, «Burp Suite(up) with fancy scanning mechanisms,» 20 Febrero 2019. [En línea]. Available: <https://www.sans.org/reading-room/whitepapers/tools/burp-suiteup-fancy-scanning-mechanisms-36557>.
- [12] P. S. Shinde y S. B. Ardhapurkar, «Cyber security analysis using vulnerability assessment and penetration testing,» de *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, Coimbatore, 2016.
- [13] P. S. Anton, R. H. Anderson, R. Mesie y M. Scheiern, The Vulnerability Assessment & Mitigation, N. D. R. Institute, Ed., p. 134.
- [14] FIRST.org, Inc, «Common Vulnerability Scoring System v3.1: Specification Document,» 2019. [En línea]. Available: <https://www.first.org/cvss/v3.1/specification-document>. [Último acceso: enero 2019].

William-Rogelio Marchand-Niño, Ingeniero de Sistemas otorgado por la Universidad Nacional del Centro del Perú, con maestría en Dirección Estratégica de TI de la Universidad de Piura, con 18 años de experiencia académica en UNAS, UDH, UPLA. Desde el año 2004 es profesor asociado en la UNAS. Ha impartido más de 90 cursos de pregrado en diferentes universidades. Instructor CISCO por 12 años. Posee múltiples certificaciones de la Industria como PMP, ITIL Foundation, CCNA, MTA. Director del Centro de Tecnologías de Información y Comunicación de la Universidad Nacional Agraria de la Selva. Miembro Senior de la IEEE.

Ana Claudia Terrones Murga, Estudiante de Ingeniería en Informática y Sistemas de la Universidad Nacional Agraria de la Selva (UNAS). Certificada Cisco CCNA R&S. Miembro del Grupo de Investigación de Redes y Seguridad de la Facultad de Ingeniería en Informática y Sistemas (FIIS-UNAS)

Analisis y Monitorización por Agrupamiento de Contenido Multimedia en Redes SDN

Jesús Antonio Puente Fernández and Luis Javier García Villalba, *Member, IEEE*

Resumen—Los proveedores actuales de contenido de vídeo requieren una gran demanda de recursos de red para obtener la mejor calidad de experiencia (QoE) en el lado del usuario. Para obtener la mejor QoE del usuario, la red debe poder anticipar problemas de ancho de banda, contenido congelado, ataques de Denegación de Servicio (DoS), entre otros. Estos problemas no pueden resolverse en las redes IP tradicionales debido a la rigidez de su arquitectura. Teniendo en cuenta los beneficios que proporciona el nuevo concepto de arquitectura de red llamado Redes definidas por software (SDN) se presenta una posible solución para resolver los desafíos de las redes existentes en términos de entrega de contenido de vídeo. Estos aspectos son posibles debido a la separación del plano de control (controlador) y el plano de datos (switches). Usando la vista centralizada de la red que proporciona SDN, es posible desarrollar una herramienta de monitorización que obtenga las métricas de la red que se utilizarán posteriormente. En este artículo se estudia la aplicación de un algoritmo de clustering que optimiza la monitorización del tráfico en las redes SDN disminuyendo el número de consultas de estadísticas para mejorar el tráfico de la red, reducir la sobrecarga y ser utilizadas para ataques DoS. Las simulaciones utilizan un algoritmo de clustering que monitoriza una transmisión de vídeo entre dos hosts demostrando la viabilidad de nuestra propuesta de monitorización.

Palabras claves—clustering, consultas estadísticas, controlador, DoS, monitorización, QoE, seguridad, SDN.

I. INTRODUCCIÓN

Hoy en día, la transmisión de vídeo es una de las aplicaciones más populares y demandadas en Internet. Requiere un gran ancho de banda y altos volúmenes de datos, lo que plantea muchos desafíos para los proveedores de servicios de Internet, que compiten por ser el mejor proveedor en entregar el tráfico de la manera más eficiente posible, al tiempo que mantienen una calidad de servicio alta y continua. La medición de la satisfacción de los usuarios finales con los servicios de red se establece como el concepto de Calidad de Experiencia (QoE).

Trabajos como [1] se centran en comprender y mejorar el concepto de QoE de la transmisión de vídeo adaptativa HTTP (HAS) actual. Esta tecnología permite adaptar la velocidad de bits de vídeo de la red en función de las demandas de transmisión. Se han presentado dos enfoques para monitorear el QoE de HAS: el primer enfoque se encuentra en el trabajo [2] y se usó como ejemplo de monitorización basado en la red [3] [4]. Las métricas de calidad de vídeo se extrajeron del tráfico de la red mediante la inspección profunda de paquetes (DPI). El segundo trabajo se enfoca en monitorizar el QoE de

manera discreta dentro de una aplicación de transmisión de vídeo de YouTube de Android (YoMoApp), proporcionando información precisa sobre los factores de QoE en cómo son percibidos por los usuarios finales [5], [6].

Las Redes Definidas por Software (SDN, del inglés Software Defined Networks) se ha convertido en un concepto para el control dinámico de la configuración de redes informáticas. Desacopla el control de red (plano de control) de los recursos de red subyacentes (plano de datos) controlados a través de un controlador centralizado. También proporciona un conjunto de interfaces de programación de aplicaciones (API) para abreviar la implementación de servicios de red como por ejemplo: enrutamiento, QoS, QoE, DoS y balanceadores de carga entre otros. La combinación de la granularidad de las estadísticas de red y la vista global de la red obtenida en el plano de datos ha creado un interés importante en la monitorización de redes usando SDN. Trabajos como [7] [8] exploran la combinación de protocolos de monitorización tradicionales como NetFlow/IPFIX y sFlow con el protocolo OpenFlow usado en redes SDN.

En este artículo se presenta una discusión de dos algoritmos de clustering que se aplican en el framework de monitorización SDN propuesto en [9]. Se utilizan para disminuir el número de peticiones de monitorización en los switches de la topología sin comprometer la precisión de los valores monitorizados, lo que permite a los administradores de red tener un estado real de los enlaces de red. Para analizar los resultados al aplicar cada algoritmo de clustering, se ha construido un escenario de prueba para transmitir un vídeo entre dos hosts (cliente y servidor) utilizando un simulador de red. Finalmente, los resultados confirman que aplicar diferentes algoritmos de clustering en la monitorización reduce el número de peticiones de estadísticas de red sin pérdida de precisión respecto a las estrategias de monitorización tradicionales.

El resto del documento se describe a continuación: El paradigma de SDN y el uso del protocolo OpenFlow se detallan en la sección II. La sección III describe los trabajos relacionados. El estudio y la discusión de los diferentes algoritmos de clustering se explica en la sección IV. Para verificar los beneficios de cada algoritmo de clustering, la sección V contiene las simulaciones y resultados de las pruebas usando dos topologías. Finalmente, la sección VI concluye con una breve discusión y conclusiones.

J. A. Puente Fernández and L. J. García Villalba. Grupo de Análisis, Seguridad y Sistemas (GASS), Departamento de Ingeniería del Software e Inteligencia Artificial, Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, España. e-mail: jesusantoniopuente@ucm.es, javiervg@fdi.ucm.es.

II. REDES DEFINIDAS POR SOFTWARE

Las redes definidas por software (SDN) son un nuevo paradigma de red que gestiona toda la red a través de un controlador de software programable centralizado. Desacopla el plano de control y el plano de datos respecto a las redes tradicionales. Usando la vista centralizada, el controlador puede administrar el tráfico de red de manera más eficiente y dinámica. Además, SDN proporciona una configuración de red dinámica adaptable a la condición de la red o los requisitos del usuario. La lógica de control es programable a través del controlador que envía paquetes de control a los switches de red utilizando el protocolo más extendido para redes SDN, que es OpenFlow [10] [11]. La arquitectura OpenFlow [10] sigue el principio SDN desacoplando el plano de datos y de control. Se basa en tres entidades principales: un switch OpenFlow (plano de datos), un controlador externo (plano de control) y el Protocolo OpenFlow [10]. Por lo tanto, el comportamiento de la red es altamente personalizable según el propósito o los requisitos del usuario. SDN se divide en tres capas lógicas:

1. **Capa de datos (recursos de red):** Está conectada a la capa superior, que es la capa de control a través de las interfaces Southbound. Estas interfaces proporcionan la abstracción de los switches programables y la conexión con el software que se ejecuta dentro del controlador (OpenFlow es el ejemplo más representativo de estas interfaces). Como se mencionó anteriormente, en estas interfaces se ejecuta un Sistema Operativo de Red (NOS) que es responsable de controlar el comportamiento de la red. Ejemplos de NOS son: NOX [12], Floodlight [13] y OpenDaylight [14] entre otros.
2. **Capa de control (controlador SDN):** Está conectada a la capa de aplicación a través de las interfaces Northbound que permiten la creación de aplicaciones y políticas de red de alto nivel que se envían al NOS. Del mismo modo, ejemplos de interfaces en dirección norte son: *Frenetic* [15], *Procera* [16], *Netcore* [17] y *McNettle* [18].
3. **Capa de aplicación:** son programas que comunican comportamientos y recursos necesarios con el controlador SDN a través de la API Northbound. Además, las aplicaciones pueden construir una vista abstracta de la red mediante la recopilación de información del controlador para la toma de decisiones.

Dado que el hardware de reenvío está separado y gestionado por el controlador, los administradores de la red pueden diseñar e implementar cualquier lógica de control utilizando el estándar OpenFlow [10], que es el más extendido en las redes SDN. De este modo, SDN proporciona flexibilidad en el desarrollo de nuevas estrategias y lógicas de red, mientras que las redes tradicionales son limitadas. Varias encuestas y trabajos sobre SDN se centran en la perspectiva de red [19]–[21] en comparación con las redes tradicionales.

OpenFlow [10] se considera uno de los primeros estándares SDN. Originalmente definió el protocolo de comunicación en entornos SDN que permite que el controlador SDN interactúe directamente con el plano de reenvío de dispositivos de red como switches y routers, tanto físicos como virtuales (basados en

hipervisores), para que pueda adaptarse mejor a los requisitos comerciales. Dicho protocolo define la comunicación entre los dispositivos de red de Capa 2 (switches) y el controlador. En una vista más profunda, proporciona la capacidad de programar las tablas de flujo ubicadas dentro de un conmutador y también puede cambiar o introducir una nueva funcionalidad en tiempo de ejecución.

III. ESTADO DEL ARTE

La monitorización de la red es el uso de un sistema que supervisa constantemente una red en busca de componentes lentos o defectuosos y lo notifica al administrador de la red en caso de interrupciones u otros problemas. La supervisión de la red es parte de la administración de la red, ya que se utiliza para recuperar una vista global del tráfico que ingresa a la red. Dichas herramientas son muy útiles para que los administradores de red analicen la red para prevenir futuros problemas de congestión, flujos de tráfico ilegítimos, entre otros problemas.

Mientras que un Sistema de Detección de Intrusos (IDS) monitoriza una red en busca de amenazas del exterior, un sistema de monitorización de red monitoriza el estado de la red en busca de problemas causados por servidores sobrecargados o bloqueados, conexiones de red u otros dispositivos que utilizan diferentes métricas entre otros. Las métricas de medición comunes son el tiempo de respuesta y la disponibilidad, aunque las métricas de consistencia y confiabilidad están comenzando a ganar gran popularidad. La incorporación generalizada de dispositivos de optimización WAN está teniendo un efecto adverso en la mayoría de las herramientas de monitorización de red, especialmente cuando se trata de medir retrasos precisos de extremo a extremo porque limitan la visibilidad del tiempo de retraso de ida y vuelta. Los fallos de solicitud de estado, como cuando no se puede establecer una conexión, tiempos de espera agotados o no se puede recuperar el documento o mensaje, generalmente producen una acción desde el sistema de monitorización.

Los métodos activos y pasivos son dos enfoques para monitorizar redes:

- **Métodos activos:** métodos que envían paquetes a dispositivos de red como switches, servidores, entre otros. Por lo tanto, dado que estos métodos inyectan paquetes en la red, impactan directamente en el tráfico de la red.
- **Métodos pasivos:** métodos compatibles con hardware de propósito especial (sniffers) o integrados en otros dispositivos como routers, switches o hosts. De forma contraria a los métodos activos, no aumentan el tráfico dentro de la red cuando miden estadísticas.

Con respecto al uso de métodos pasivos en redes tradicionales, los protocolos como el Protocolo de administración de red (SNMP) [22] y el Protocolo de configuración de red (NETCONF) [23] permiten monitorizar los dispositivos y las estadísticas de la red. De manera similar, las herramientas de monitorización como NetFlow [24] y sFlow [25] se usan para estimar tanto una muestra como estadísticas completas de tráfico en redes basadas en flujo. jFlow [26] es una extensión del lenguaje Java [27] que permite verificar las anotaciones de

flujo de forma estática. Además, proporciona un compilador que verifica (también estáticamente) y evita fugas de información a través de canales de almacenamiento. jFlow proporciona características que hacen que la comprobación de flujo no sea tan restrictiva como otros lenguajes de programación.

Los métodos pasivos y activos para medir el rendimiento de la red se utilizan en [28]. Utiliza beacons para enviar paquetes sonda e instalar flujos adicionales en los switches. Luego, estos beacons se utilizan para estimar la tasa de pérdida de paquetes y el retraso. Una solución híbrida entre métodos pasivos y activos es el framework propuesto en [29]. Este trabajo define un framework de monitorización de red basado en un módulo orquestador con un método flexible para recuperar estadísticas de red.

OpenNetMon [30] monitoriza las métricas por flujo en las redes OpenFlow. En lugar de utilizar una estrategia única para la monitorización, la aplicación de diferentes métodos para consultar estadísticas puede ayudar a reducir la sobrecarga en los switches y en las redes. En este orden, OpenTm [31] propone seguir una distribución de consulta no uniforme con respecto a un esquema uniforme. Demuestra que esta estrategia es mucho más rápida que las formas existentes de estimación del tráfico en las redes IP. Chowdhury et al. [32] propone Payless, un framework de monitorización eficiente de bajo coste de redes SDN. Su característica principal es que proporciona una vista abstracta de la red y una manera regular de solicitar estadísticas sobre los recursos de la red. Además, dado que se ha desarrollado como un conjunto de componentes conectables, proporciona interfaces para conectarlos a todos (API RESTful de alto nivel).

Continuando con la monitorización de redes de flujo, Flowsense [33] es un enfoque que busca estadísticas de alta precisión con un coste cero de medición utilizando la separación física del plano de control y de datos en SDN. Las características clave son: la duración que tarda el flujo en la entrada de la tabla de flujo, la cantidad de tráfico que coincide con ese flujo y, finalmente, el puerto de entrada del tráfico que coincide con la entrada. Sin embargo, tiene algunas limitaciones, ya que depende en gran medida del tipo de tráfico que va a monitorizar, ya que los flujos grandes como una transmisión de vídeo pueden retrasar su cálculo y su utilización.

Suárez-Varela et al [34] propone un sistema de monitorización habilitado para SDN y reconocimiento de aplicaciones. Este sistema de monitorización está dividido en dos bloques. El primer bloque, llamado SBAR, es un sistema basado en OpenFlow que produce informes de monitorización a nivel de flujo. El segundo bloque es una herramienta de análisis de datos que procesa los informes resultantes de SBAR y muestra estadísticas valiosas de red a través de una interfaz gráfica de usuario web. SBAR, a su vez, está compuesto por un módulo de medición que se encarga de mantener las mediciones de flujo en los switches e informarlas a los controladores, y un módulo de clasificación, que produce etiquetas de clasificación para cada flujo en los informes de medición.

IV. ESTUDIO DE ALGORITMOS DE CLUSTERING PARA MONITORIZAR REDES SDN

Clustering es la tarea de agrupar un conjunto de objetos de tal manera que los objetos en el mismo grupo (llamado cluster) son lo más similares (en cierto sentido) entre sí que con los de otros grupos. Es una tarea principal de la minería de datos exploratoria, y una técnica común para el análisis estadístico de datos, utilizado en muchos campos, incluido el aprendizaje automático, el reconocimiento de patrones, el análisis de imágenes, la recuperación de información, la bioinformática, la compresión de datos y los gráficos por ordenador.

El análisis de Clustering en sí no es un algoritmo específico, sino que es la tarea general a resolver. Se puede lograr mediante varios algoritmos que difieren significativamente en su comprensión de lo que constituye un cluster y cómo encontrarlos de manera eficiente. Las nociones populares de clusters incluyen clusters con pequeñas distancias entre los miembros del grupo, áreas densas del espacio de datos, intervalos o distribuciones estadísticas particulares entre otros. Por lo tanto, el término clustering puede formularse como un problema de optimización de objetivos múltiples. El algoritmo de clustering apropiado y la configuración de parámetros (incluidos los parámetros como la función de distancia a utilizar, un umbral de densidad o el número de clusters esperados) dependen del conjunto de datos individual y del uso previsto de los resultados. El análisis de conglomerados como tal no es una tarea automática, sino un proceso iterativo de descubrimiento de conocimiento u optimización interactiva de objetivos múltiples que implica pruebas y fallos. A menudo es necesario modificar el procesamiento previo de datos y los parámetros del modelo hasta que el resultado alcance las propiedades deseadas. Resumiendo, el objetivo de usar la agrupación es maximizar la variación inter-clusters y minimizar la variación intra-clusters.

Hemos utilizado dos algoritmos de agrupación en este trabajo:

- **Kmeans** [35]: es probablemente el algoritmo de clustering más conocido y el más simple entre los algoritmos de aprendizaje no supervisados. Este trabajo funciona según el principio de agrupación Kmeans. En realidad, esto significa que los grupos (clusters) para un conjunto de datos dado, están representados por una variable “k”. Para cada grupo, se define un centroide. El centroide es un punto de datos presente en el centro de cada grupo (considerando la distancia euclidiana).
- **Expectation-Maximization (EM)** [36]: es un método iterativo para buscar la máxima probabilidad o el máximo. Este algoritmo se basa en la distribución gaussiana en estadística. Considera una colección de distribuciones gaussianas para el conjunto de datos en un problema de LA.

Este trabajo es una extensión de la prueba de concepto utilizada en [9] donde las diferencias entre clustering (usando el algoritmo kmeans) y el algoritmo optimizado de conservación de flujo se pusieron de manifiesto usando la tasa de datos como métrica única en las simulaciones. Entonces, tomando como referencia los resultados de [9] (usando el algoritmo kmeans),

se han extendido las pruebas usando el algoritmo EM para comparar los resultados entre ambos algoritmos.

El presente trabajo se basa en la aplicabilidad y el estudio de los beneficios al aplicar diferentes algoritmos de clustering en la herramienta de monitorización presentada en [9]. Este estudio se verifica utilizando las métricas de tasa de datos y tasa de error, que también se describen en [9], para afirmar que es posible disminuir el número de peticiones de monitorización mientras se mantiene la precisión de los datos monitorizados.

Por tanto, el objetivo del presente trabajo es optimizar el número de solicitudes de switches utilizando diferentes algoritmos de clustering para reducir el número de solicitudes de switches respecto a técnicas no optimizadas entre las que se encuentran las herramientas de monitorización todos los switches de una topología. Es importante aclarar que los algoritmos de agrupación no se utilizan para optimizar el número de peticiones, ya que se utilizan para crear clusters sobre los cuales se ejecutan las técnicas de optimización descritas en [9]. Después de esto, se continúa aplicando diferentes algoritmos, obteniendo diferentes asignaciones de clusters y, por lo tanto, diferentes resultados en las simulaciones.

Antes de empezar a realizar peticiones de estadísticas a los switches, la función $Clustering_{Function}(G)$ lee la topología de red estructurada lógicamente en un grafo $G = (S, L)$ el cual será usado para construir los clusters. Debido a que G contiene el grafo de la red, la función $Create_Cluster(G, Algorithm, nc)$ los divide en N clusters, nombrándolos como $c_0, c_1, \dots, c_{nc-1}$.

La agrupación de los switches se basa en el número de puertos que están compuestos y la importancia de aplicar uno de los algoritmos implementados: Kmeans o Expectation-Maximization descritos en la sección anterior.

Una vez que se ha seleccionado el algoritmo, los criterios para monitorizar los clusters dependen del número de puertos de los switches. Por un lado, la herramienta de monitorización aplica el principio del algoritmo de conservación de flujo en switches de 2 grados, ya que todo el tráfico de datos que recibe el puerto entrante del switch se reenviará al puerto saliente del mismo switch y, por lo tanto, el tráfico a través de dichos enlaces será casi el mismo valor. Por otro lado, los clusters que contienen switches con tres o más puertos se eligen aleatoriamente en cada período T_{mon} que se supervisará. Para más nivel de detalle, estos algoritmos se describen en las secciones **4.2 Flow-Conservation Algorithm Enhancement** y **4.3 Clustering Enhancement Technique** de [9].

Para propósitos experimentales, hemos elegido la tasa de datos y la tasa de error como métricas para demostrar la viabilidad de nuestra propuesta. El procedimiento para calcular ambas métricas se describe en la sección **4.3 Clustering Enhancement Technique** de [9]. Concretamente, los procedimientos para calcular la tasa de datos y las métricas de error se explican en detalle en el Algoritmo 2 y el Algoritmo 4, respectivamente.

V. SIMULACIONES Y RESULTADOS

En esta sección, probamos los beneficios de los algoritmos de Clustering utilizando la herramienta de monitorización de optimización de Clustering descrita en la sección **4 Enhanced Monitored Algorithm** de [9].

Hemos realizado las pruebas en el escenario descrito en la Figura 1 atendiendo al uso de la optimización del clustering utilizando una plataforma de simulación llamada Mininet v2.1.0 [21]. La topología se compone de 7 OF-Switches ($s_1, s_2, s_3, s_4, s_5, s_6, s_7$) y 2 host (h_1 y h_2) conectados a s_1 y s_7 respectivamente. Los enlaces [$switch : s_3 - puerto : p_4, switch : s_6 - puerto : p_1$] y [$switch : s_6 - port : p_2, switch : s_5 - port : p_3$] están configurados con valores de velocidad de datos máxima (1 Mbps de ancho de banda) y porcentaje de pérdida del 5%.

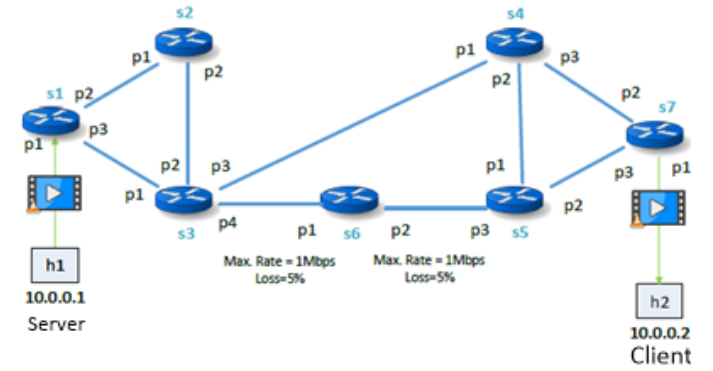


Figura 1. Topology tested.

Una simulación consiste en una transmisión de vídeo enviada desde el host 1 al host 2 usando el servidor de vídeo VLC y RTP/UDP como protocolo de transmisión al mismo tiempo que se ejecuta el módulo de monitorización. Por lo tanto, la tarea principal del módulo de monitorización es medir las estadísticas de los switches de la red (tanto los enlaces donde se envía el vídeo como el resto de ellos). El vídeo utilizado en la transmisión de vídeo es "Highway_cif" [38] y el tiempo de monitorización (t_{mon}) es de 200 ms.

La construcción de clusters será diferente atendiendo a la aplicación de los algoritmos de clúster Kmeans y EM como muestra la Tabla I. Una vez que se han creado los clusters, la herramienta de monitorización aplica las diferentes estrategias de monitorización descritas en [9].

Tabla I
CLUSTER ASSIGNMENT IN TEST SCENARIO

Clusters	Kmeans	EM
Cluster 0 (C_0)	S_3	$S_4S_5S_6$
Cluster 1 (C_1)	S_1S_6	S_1
Cluster 2 (C_2)	S_4S_5	S_2S_7
Cluster 3 (C_3)	S_2S_7	S_3

Como se ha indicado anteriormente, los resultados de las simulaciones aplicando el algoritmo kmeans se exponen en [9]. El número de consultas de monitorización en el método no optimizado durante la simulación (80 segundos) fueron 2079 solicitudes, mientras que en el algoritmo optimizado, la

aplicación del algoritmo de clustering Kmeans fue 903. Supuso una relación de reducción del 57 % en solicitudes con respecto a una simulación no optimizada.

Comenzando con los resultados de la simulación realizados en este trabajo y tomando la tasa de datos como la medición que aplica el algoritmo EM, el número de consultas de monitorización es 888. La diferencia de 1191 consultas (con el algoritmo no optimizado que fueron 2079 peticiones) muestra un relación de reducción del 58 %.

La Figura 2 muestra el flujo de tráfico entre dos simulaciones en las cuales una aplica la optimización usando el algoritmo EM respecto a otra que no lo aplica. La Figura describe el flujo de datos (en bps) a través de los enlaces $s_3 - s_6$. Las líneas puntiagudas (líneas azules) muestran la tasa de datos que el servidor está enviando, mientras que las líneas continuas (líneas verdes) muestran la tasa de datos obtenida utilizando el método no optimizado y las líneas de puntos (líneas rojas) muestran la velocidad de datos con algoritmo optimizado. Como se esperaba, los enlaces experimentan un aumento de la velocidad de datos debido a la transmisión del vídeo entre h1 y h2. Tan pronto como finaliza la transmisión (alrededor de 200 solicitudes), la tasa de datos disminuye, lo que demuestra la eficiencia del algoritmo para detectar cambios en la transmisión de la red. Estos resultados confirman que la optimización mantiene buenos niveles de precisión y reduce el número de solicitudes en el plano de datos utilizando dos criterios diferentes para construir el conjunto de switches. La diferencia entre el enlace $s_3 - s_6$ aplicando y sin aplicar la optimización es de 41,22 Kbps.

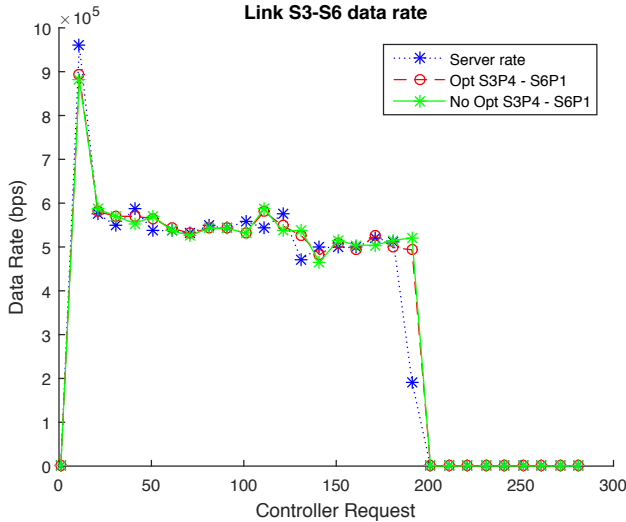


Figura 2. Data rate of link s_3-s_6

Por otro lado, la Figura 3 muestra la tasa de error en los datos de tráfico a través de la ruta de transmisión, es decir el enlace $s_3 - s_6$ (*switch* : $s_3 - puerto$: p_4 , *switch* : $s_6 - puerto$: p_1). Las líneas continuas (líneas verdes) muestran la tasa de error obtenida utilizando el método no optimizado, las líneas punteadas (líneas rojas) muestran la tasa de error con el algoritmo optimizado y la línea continua punteada (líneas azules) muestra el porcentaje netem configurado en Mininet

(5 % de pérdida de paquetes). Como se esperaba, los enlaces pueden detectar la pérdida de información entre los enlaces debido a las características de estos enlaces.

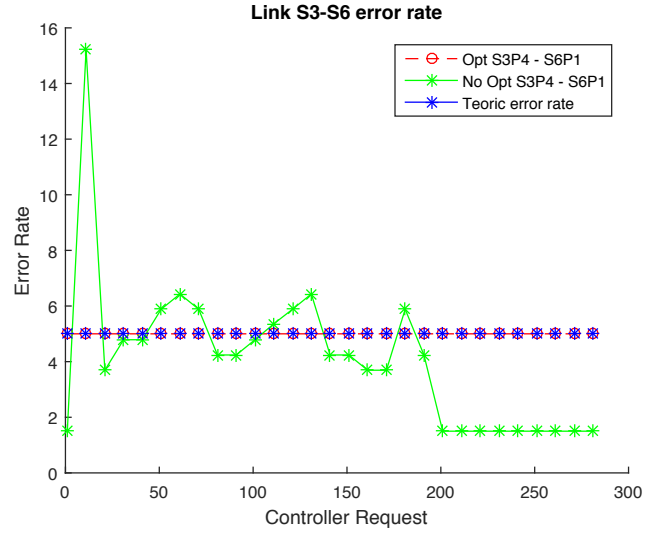


Figura 3. Error rate of link s_3-s_6

VI. CONCLUSIONES

SDN ha traído muchas oportunidades en el ámbito de las redes que han permitido la adaptación dinámica y la reconfiguración con su separación del plano de control del dispositivo de reenvío controlado por el controlador centralizado. El controlador SDN gestiona los flujos de red de forma dinámica e individual con la visión global de toda la red. El surgimiento de SDN con su capacidad de control dinámica puede traer varios beneficios a centros de datos en la nube donde, en esencia, los requisitos y la utilización cambia dinámicamente bajo demanda. Hay una amplia gama de algoritmos de clustering, cada uno con sus ventajas e inconvenientes con respecto a qué tipo de datos agrupan, complejidad de tiempo, debilidades, etc. Existen otros algoritmos de clustering entre los que se encuentran por ejemplo clustering aglomerativo jerárquico (o clustering de vinculación), bueno para cuando no haya necesariamente grupos circulares (o hiper esféricos), y no conozca la cantidad de grupos por adelantado. Comienza con cada punto siendo un cluster separado y funciona uniendo los dos grupos más cercanos en cada paso hasta que todo esté en un grupo grande. Este artículo presenta un estudio de los beneficios aplicando diferentes algoritmos de clustering en una herramienta de monitorización que reduce la cantidad de solicitudes de monitorización en los dispositivos que la red que está compuesto. Esta reducción en las consultas a los switches de la red depende de la construcción de los clusters. Hemos demostrado que usando diferentes algoritmos de clustering para construir diferentes conjuntos/grupos de switches, proporciona beneficios disminuyendo el número de consultas de monitorización en los switches de la topología. Usando los valores de tasa de datos y tasa de error obtenidos a partir de los resultados de las simulaciones, se confirman

los beneficios de los algoritmos de clustering respecto a las simulaciones no optimizadas.

ACKNOWLEDGMENT

This paper has also received funding from THEIA (Techniques for Integrity and authentication of multimedia files of mobile devices) UCM project (FEI-EU-19-04).

REFERENCIAS

- [1] M. T. Seufert, "Quality of Experience and Access Network Traffic Management of HTTP Adaptive Video Streaming", Doctoral Thesis, University of Würzburg, 2017. [Online]. Available: https://opus.bibliothek.uni-wuerzburg.de/files/15413/Seufert_Michael_Thomas_HTTP.pdf
- [2] P. Casas, M. Seufert, and R. Schatz, "YOUQMON: A System for Online Monitoring of YouTube QoE in Operational 3G Networks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 2, pp. 44–46, 2013.
- [3] M. Seufert, M. Slanina, S. Egger, and M. Kottkamp, "To Pool or not to Pool: A Comparison of Temporal Pooling Methods for HTTP Adaptive Video Streaming," in *Proceedings of the 5th International Workshop on Quality of Multimedia Experience (QoMEX)*, Klagenfurt, Austria, 2013.
- [4] M. Wennesheimer, D. Robinson, J. De Vriendt, D. De Vleeschauwer, C. Bahr, K. Heise, S. Argyropoulos, B. Feiten, M. Bilgic, T. Raiha, S. Egger, M. Seufert, M. Slanina, M. Kottkamp, H. Gsödl, H. Ibl, and R. García Pérez, "Service Quality Definition and Measurement," *Next Generation Mobile Networks Alliance (NGMN)*, Tech. Rep., 2013, version 1.0.4
- [5] F. Wamser, M. Seufert, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "YoMoApp: a Tool for Analyzing QoE of YouTube HTTP Adaptive Streaming in Mobile Networks," in *Proceedings of the European Conference on Networks and Communications (EuCNC)*, Paris, France, 2015.
- [6] M. Seufert, F. Wamser, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "YouTube QoE on Mobile Devices: Subjective Analysis of Classical vs. Adaptive Video Streaming," in *Proceedings of the 6th International Workshop on Traffic Analysis and Characterization (TRAC)*, Dubrovnik, Croatia, 2015.
- [7] K. Giotis, C. Argyropoulos, G. Androulidakis, D. Kalogeras, V. Maglaris, Combining OpenFlow and sFlow for an Effective and Scalable Anomaly Detection and Mitigation Mechanism on SDN Environments, *Computer Networks* 62 (1) (2013) 122–136.
- [8] A. Zaalouk, R. Khondoker, R. Marx, K. M. Bayarou, Orchsec: An orchestrator-based architecture for enhancing network-security using network monitoring and sdn control functions., in: *NOMS*, 2014, pp. 1–9.
- [9] Puente Fernández, J., García Villalba, L., and Kim, T. H. (2018). Clustering and Flow Conservation Monitoring Tool for Software Defined Networks. *Sensors*, 18(4), 1079.
- [10] McKeown N.; Anderson T.; Balakrishnan H.; Parulkar G.; Peterson L.; Rexford J.; Shenker S.; Turner J. OpenFlow: Enabling Innovation in Campus Networks. *ACM SIGCOMM Computer Communication Review*; 2008; vol. 38 (2); pp. 69–74.
- [11] OpenFlow Switch Specification v1.1.0. Available at <http://archive.openflow.org/documents/openflow-spec-v1.1.0.pdf>
- [12] Gude N.; Koponen T.; Pettit J.; Pfaff B.; Casado M.; McKeown N.; Shenker S. NOX: Towards an Operating System for Networks. *ACM SIGCOMM Computer Communication Review*; 2008; vol. 38; pp. 105–110.
- [13] Project Floodlight: Open Source Software for Building Software-Defined Networks. Available at <http://www.projectfloodlight.org>
- [14] OpenDaylight (ODL) open source SDN platform. Available at <https://www.opendaylight.org>
- [15] Foster, N.; Harrison R.; Freedman M. J.; Monsanto C.; Rexford J.; Story A.; Walker, D. Frenetic: A Network Programming Language. In *ACM SIGPLAN Notices*; 2011; vol. 46(9); pp. 279–291.
- [16] Voellmy A.; Kim H.; Feamster N. Procer: a Language for High-level Reactive Network Control. In *Proceedings of the first workshop on Hot topics in software defined networks*; ACM; 2012; pp. 43–48.
- [17] Monsanto C.; Foster N.; Harrison R.; Walker D. A Compiler and Runtime System for Network Programming Languages. In *ACM SIGPLAN Notices*; ACM; 2012; vol 47(1); pp. 217–230.
- [18] Voellmy A.; Wang J. Scalable Software Defined Network Controllers. In *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*; ACM; 2012; pp. 289–290.
- [19] M. Casado, N. Foster, A. Guha, Abstractions for software-defined networks, *Communications of the ACM*; 2014; pp. 86–95.
- [20] N. Feamster, J. Rexford, E. Zegura, The Road to SDN: An Intellectual, History of Programmable Networks, *ACM SIGCOMM Computer Communication Review* 44 (2) (2014) 87–98.
- [21] V. K. Gurbani, M. Scharf, T. Lakshman, V. Hilt, E. Marocco, Abstracting network state in software defined networks (sdn) for rendezvous services, in: *2012 IEEE International Conference on Communications (ICC)*, IEEE, 2012, pp. 6627–6632
- [22] Case J. D.; Fedor M.; Schoffstall M. L.; Davin J. Simple Network Management Protocol (SNMP). RFC 1157 (Historic); 1990.
- [23] Enns R.; Bjorklund M.; Schoenwaelder J.; Bierman A. Network Configuration Protocol.
- [24] Claise B. RFC 3954 - Cisco Systems NetFlow Services Export Version 9. RFC 3954; 2004.
- [25] Phaal P.; Lavine M. Sflow version 5. 2004.
- [26] Myers A. C. JFlow: Practical Mostly-static Information Flow Control. In *Proceedings of the 26th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*; ACM; 1999; pp. 228–241.
- [27] Gosling, J., Joy, B., Steele, G. L., Bracha, G., and Buckley, A. (2014). *The Java Language Specification*. Pearson Education.
- [28] Shibuya M.; Tachibana A.; Hasegawa T. Efficient Performance Diagnosis in OpenFlow Networks Based on Active Measurements. 1st ACM Conference on Information-Centric Networking (ICN-2014); 2014; pp. 279.
- [29] Valdivieso Caraguay A. L.; Puente Fernández J. A.; García Villalba L. J. An Optimization Framework for Monitoring of SDN/OpenFlow Networks. *International Journal of Ad Hoc and Ubiquitous Computing (IJAHUC)*; 2015.
- [30] Van Adrichem N. L.; Doerr C.; Kuipers F. A. Opennetmon: Network Monitoring in Openflow Software-defined Networks. In *Network Operations and Management Symposium (NOMS)*; IEEE; 2014; pp. 1–8.
- [31] Tootoonchian A.; Ghobadi M.; Ganjali Y. OpenTM: Traffic Matrix Estimator for OpenFlow Networks. In *Passive and active measurement*; 2010; pp. 201–210.
- [32] Chowdhury, S. R., Bari, M. F., Ahmed, R., and Boutaba, R. (2014, May). Payless: A low cost network monitoring framework for software defined networks. In *2014 IEEE Network Operations and Management Symposium (NOMS)* (pp. 1–9). IEEE.
- [33] Yu, C., Lumezanu, C., Zhang, Y., Singh, V., Jiang, G., Madhyastha, H. V. (2013, March). Flowsense: Monitoring network utilization with zero measurement cost. In *International Conference on Passive and Active Network Measurement* (pp. 31–41). Springer Berlin Heidelberg.
- [34] J. Suárez-Varela, P. Barlet-Ros, Sbar: Sdn flow-based monitoring and application recognition, in: *Proceedings of the Symposium on SDN Research*, ACM, 2018, p. 22
- [35] HARTIGAN, John A.; WONG, Manchek A. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1979, vol. 28, no 1, p. 100–108.
- [36] MOON, Todd K. The expectation-maximization algorithm. *IEEE Signal processing magazine*, 1996, vol. 13, no 6, p. 47–60.
- [37] Mininet. Available at <https://github.com/mininet/mininet/wiki/Introduction-to-Mininet>
- [38] Highway. Available at <http://www2.tkn-berlin.de/research/evalvid/qcif.html>

Jesús Antonio Puente Fernández was born in Madrid (Spain) in 1988. Jesús received his Computer Science Engineering degree by Universidad Complutense de Madrid (Spain) in 2012. He received a M.S. degree in Computer Research in the Universidad Complutense de Madrid (Spain) in 2014. He is currently a Ph.D. student of Computer Engineering in Universidad Complutense de Madrid (Spain). His research interests include computer networks, software-defined networking and network function virtualization. In addition, he is currently working as IT security evaluator and Project manager in DEKRA Testing and Certification.

Luis Javier García Villalba received a Telecommunication Engineering degree from the Universidad de Málaga (Spain) in 1993 and holds a Ph.D. in Computer Science (1999) from the Universidad Politécnica de Madrid (Spain). Visiting Scholar at COSIC (Computer Security and Industrial Cryptography, Department of Electrical Engineering, Faculty of Engineering, Katholieke Universiteit Leuven, Belgium) in 2000 and Visiting Scientist at IBM Research Division (IBM Almaden Research Center, San Jose, CA, USA) in 2001 and 2002, he is currently Associate Professor of the Department of Software Engineering and Artificial Intelligence at the Universidad Complutense de Madrid (UCM) and Head of Complutense Research Group GASS (Group of Analysis, Security and Systems) which is located in the Faculty of Computer Science and Engineering at the UCM Campus. His professional experience

includes the management of both national and international research projects and both public (Spanish Ministry of R&D, Spanish Ministry of Defence, Horizon 2020 - European Commission, ...) and private financing (Hitachi, IBM, Nokia, Safelayer Secure Communications, TB Solutions Security, ...). Author or co-author of numerous international publications is editor or guest editor of numerous journals such as Entropy MPDI, Future Generation Computer Systems (FGCS), Future Internet MDPI, IEEE Latin America Transactions, IET Communications (IET-COM), IET Networks (IET-NET), IET Wireless Sensor Systems (IET-WSS), International Journal of Ad Hoc and Ubiquitous Computing (IJAHUC), International Journal of Multimedia and Ubiquitous Engineering (IJMUE), Journal of Supercomputing, Sensors MDPI, etc.

Encryption with TLS Protocol version 1.2 and Web Sites Performance. A Case Study.

William-Rogelio Marchand-Niño¹
william.marchand@unas.edu.pe

Edgar Etson Rueda Liberato¹
edgar.rueda@unas.edu.pe

Resumen— El protocolo TLS ofrece una comunicación web segura porque la transmisión de información está encriptada entre el cliente y el servidor, gracias al intercambio de claves simétricas y asimétricas, garantizando la confidencialidad e integridad de los datos. El uso de algoritmos de encriptación robustos y el intercambio de claves son la base para la generación del canal encriptado utilizado en las conexiones web, conocido como HTTPS, sin embargo, esto agrega el consumo de recursos computacionales que afectan el tiempo de procesamiento, la velocidad y el número de conexiones simultáneas. Estas son una medida del rendimiento de los sitios web. Para medir el impacto del Protocolo TLS versión 1.2, se utilizaron cinco sitios web de producción de una empresa de desarrollo de aplicaciones y sitios web, donde se realizó la evaluación del rendimiento del sitio web

Palabras clave— SSL/TLS, TLS, rendimiento web, autoridad de certificación, HTTPS, seguridad web.

I. INTRODUCCIÓN

El protocolo SSL/TLS fue creado por la empresa NetScape en el año 1994, inicialmente fue denominado SSL (Secure Socket Layer) [1] y posteriormente por medio de la IETF [2] y las mejoras realizadas a dicho protocolo paso a ser TLS (Transport Layer Security) [3]. Actualmente el término más usado es SSL cuando se refiere al protocolo SSL/TLS. El trabajo de investigación está enfocado al protocolo HTTPS, que es el protocolo HTTP sobre SSL/TLS [4].

Según el Informe de Telemetría del año 2017 [5] elaborado por David Holmes y publicado en el mes de abril del 2018, detalla que más del 80% de las páginas web a nivel mundial están haciendo uso del protocolo SSL/TLS.

Actualmente TLS en su versión 1.2 es la más utilizada en comparación con sus antecesoras y con la última versión 1.3 definida en agosto de 2018. Según el reporte por “Qualys SSL Labs” del mes de abril de 2019 [6], existe un 95% de sitios web seguros con soporte para TLS versión 1.2 sobre una base de 150,000 sitios web (ver Fig. 1), es así, como este protocolo desarrollado por la empresa Netscape Communications e implementado en sus inicios por Netscape Navigator versión 1.1 llega a ser uno de los más importantes en cuanto al acceso seguro a la Wolrd Wide Web.

Cada vez son más los administradores que configuran los certificados digitales de pago o gratuitas emitidos por Autoridades Certificadoras como Let's Encrypt, Godaddy, Start SSL que son algunas de las conocidas en el mercado.

Sin embargo, el aplicar un algoritmo de cifrado a cierta información hace que se requiera un mayor consumo de recursos computacionales en comparación que no se aplique. Esto debido a que la información pasa por un proceso computacional donde el texto legible será transformado a texto sin sentido, además de establecer un proceso de negociación

entre extremos para determinar los mecanismos y algoritmos a utilizar.

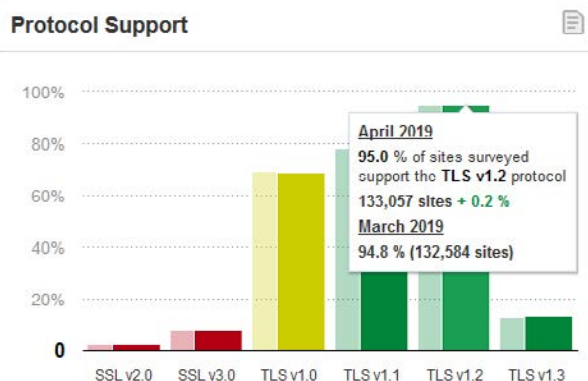


Fig. 1. Porcentaje de sitios web con soporte del protocolo TLS v.1.2 (SSLA Labs - Qualys)

Esta investigación se enfoca en el Protocolo Seguro de Transferencia de Hipertexto (HTTPS), que permite verificar la autenticidad de una página web por medio de un certificado digital. Este protocolo también hace uso de algoritmos de cifrado simétrico y asimétrico en el proceso de intercambio de claves entre navegador y servidor para establecer un canal seguro para la transferencia de la información.

Actualmente el uso de SSL/TLS con el protocolo HTTP es cada vez más importante, debido a las constantes vulnerabilidades asociadas a las comunicaciones e intercambio de datos por Internet; esas vulnerabilidades pueden ser aprovechadas por los actores de amenazas bajo diversas formas de amenazas, y una de ellas es la interceptación de los mensajes intercambiados entre los servicios web, y si estás viajan por los medios de transmisión sin la protección adecuada (cifrado), el riesgo es alto contra la confidencialidad por ejemplo.

El uso de nuevas versiones del protocolo SSL/TLS, específicamente TLS, permite disponer de suites de cifrado más complejas para proporcionar mayor seguridad. Muchas veces las organizaciones no cuentan con los recursos computacionales o buenas prácticas en optimización de los sitios web para que estas sean más ligeras (expresado en Megabytes del sitio web) y a la vez más rápidas ante

peticiones en grandes volúmenes.

A nivel técnico, el análisis del uso del protocolo SSL/TLS y de la suite de cifrado en su versión TLS 1.2 en los sitios web, permite entender si este genera un impacto en el rendimiento

¹ Grupo de Investigación en Redes, Seguridad y Gestión de TI
Universidad Nacional Agraria de la Selva - Tingo María, Perú

de los sitios web, considerando la confidencialidad, integridad y autenticación de la información.

El estudio de este protocolo permite a los administradores web y a los usuarios finales, quienes hacen uso de un navegador web, comprender la importancia y la influencia que genera el uso del protocolo SSL/TLS, específicamente TLS v1.2 cuando se navega en Internet, con esto se pretende generar confianza tanto para los administradores web respecto al nivel de rendimiento y seguridad informática, como también para los usuarios consumidores de los sitios web respecto a la navegación confiable y segura en Internet.

La evaluación del impacto generado en el rendimiento de los sitios web del caso de estudio permite demostrar su impacto en el rendimiento, consolidando la aplicación de buenas prácticas en cuanto al uso y adecuada configuración del protocolo TLS v.1.2 para la organización y los clientes que administran sus sitios web. Otro factor adicional es, validar las recomendaciones realizadas por los fabricantes y marcas de la industria respecto a SSL/TLS.

Ante este escenario, la interrogante planteada es ¿Cuál es el impacto del cifrado con el protocolo TLS v1.2 en el rendimiento de sitios web? Para responder a la interrogante se tiene en un escenario de pruebas para la medición del rendimiento de cinco sitios web en producción alojados en un Servidor Web de la empresa Web-Out S.A. como parte del caso de estudio. Por el lado del cliente en estas pruebas se usó Windows 10 como anfitrión y máquinas virtuales con Centos 7 minimalista y Ubuntu 19; para la virtualización se ha empleado la herramienta Virtual Box en su versión 6.0.6.

II. ESCENARIO DE PRUEBAS

A. Objetivo

Evaluar el impacto del cifrado con el protocolo SSL/TLS en TLS versión 1.2 en el rendimiento de los sitios web. Caso empresa Web-Out S.A.

B. Escenario para las pruebas

El ambiente que se ha utilizado se expresa en la Fig. 2, con las características de los VPS (Servidor Virtual Privado) mostradas en la Tabla I y II.

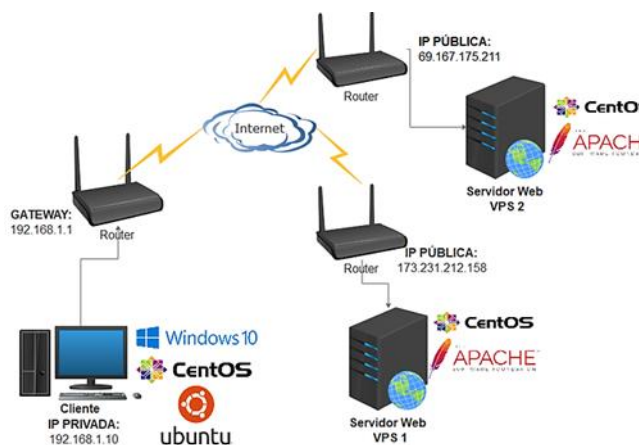


Fig. 2. Topología del caso de estudio, indicando distribución y direcciones IP.

TABLA I. CARACTERÍSTICAS TÉCNICAS DE HARDWARE Y SOFTWARE DEL SERVIDOR WEB (VPS 1)

CENTOS 7.6 MINIMALISTA (VPS 1)	HARDWARE	
	Disco Duro	150 Gb - SSD
	Memoria RAM	6 Gb
	Procesador	48 Procesadores
	Dirección IP Pública	173.231.212.158
	Ancho de banda	5 TB
	Número de núcleos	12
	Alta Disponibilidad	SI
	SOFTWARE	
	Servidor Web	Apache 2.4.39
	Modulo SSL/TLS	mod_ssl
	Criptografía	OpenSSL versión 1.0.2k-
	PHP	versión 7.1.28
	Mysql	10.2.24-MariaDB

TABLA II. CARACTERÍSTICAS TÉCNICAS DE HARDWARE Y SOFTWARE DEL SERVIDOR WEB (VPS 2)

CENTOS 7.6 MINIMALISTA (VPS 2)	HARDWARE	
	Disco Duro	2 Gb - HDD
	Memoria RAM	4 Gb
	Procesador	48 Procesadores
	Dirección IP Pública	69.167.175.211
	Ancho de banda	40 GB
	SOFTWARE	
	Servidor Web	Apache 2.4.39
	Modulo SSL/TLS	mod_ssl
	Criptografía	OpenSSL versión 1.0.2k-
	PHP	versión 5.6
	Mysql	5.6

Se recopiló la información de los cinco sitios web, como dirección IP pública, puerto empleado en la conexión por HTTPS, protocolos habilitados y deshabilitados en el servidor para cada sitio web que se muestran en la Tabla IV empleando la herramienta SSLRobot.

Los cinco sitios web para las pruebas del estudio han sido desarrollados utilizando el CMS (Sistema de Gestión de Contenidos) Drupal versión 7 y 8, conteniendo código HTML (Lenguaje de Marcado de Hipertexto), CSS (Hojas de Estilo en Cascada) y JavaScript como lenguaje de programación. Los cinco sitios web están alojados en dos VPS así como se detalla en la Tabla III

TABLA III. SITIOS WEB, URL DE ACCESO Y UBICACIÓN EN EL VPS

Sitio Web	URL	VPS
Web-Out S.A.	www.web-out.com	VPS 1
Facultad de Ciencias Económicas y Administrativas de la UNAS	www.fceauas.edu.pe	VPS 1
Hotel Oro Verde	www.hotel-oroverde.com	VPS 2
Hotel Natural Green	www.hotelnaturalgreen.com	VPS 1
Cámara de Comercio Canadá - Perú	www.canadaperu.org	VPS 1

TABLA IV. INFORMACIÓN DEL PROTOCOLO SSL/TLS DE LOS CINCO SITIOS WEB

	Web-Out	Facultad de Ciencias Económicas y Administrativas de la UNAS	Hotel Oro Verde	Hotel Natural Green	Cámara de Comercio Canadá – Perú
Servidor Virtual Privado	VPS 1	VPS 1	VPS 2	VPS 1	VPS 1
IP pública	173.231.212.158	173.231.212.158	69.167.175.211	173.231.212.158	173.231.212.158
Puerto	443	443	443	443	443
Protocolos Habilitados	TLS 1.2 TLS 1.1 TLS 1.0	TLS 1.2 TLS 1.1 TLS 1.0	TLS 1.2 TLS 1.1 TLS 1.0	TLS 1.2 TLS 1.1 TLS 1.0	TLS 1.2 TLS 1.1 TLS 1.0
Protocolos Deshabilitados	SSL 3.0 SSL 2.0	SSL 3.0 SSL 2.0	SSL 3.0 SSL 2.0	SSL 3.0 SSL 2.0	SSL 3.0 SSL 2.0

Al instalar el mod_ssl el servidor Web posee un nuevo fichero de nombre ssl.conf en las que se configura el nombre de dominio, dirección local del certificado digital que ha sido auto-firmado por medio de la herramienta OpenSSL [7]

III. ANTECEDENTES Y REVISIÓN BIBLIOGRÁFICA

Existen diversas investigaciones que se enfocan en la implementación del protocolo SSL/TLS en las comunicaciones entre cliente y servidor, en las que se realizaron un análisis en servidores web con una configuración del modSSL en Apache para la entrega segura por medio de TLS, obteniendo como resultado que los cálculos RSA son la operación más costosa en TLS hablando en términos de rendimiento, ya que consume entre un 13% a 58% del tiempo pasado en el servidor web. También mencionan que a medida que el rendimiento de las CPU continúa creciendo, la sobrecarga de TLS disminuirá. Es por eso por lo que invertir en CPU más rápidas o adicionales parece ser una estrategia preferible para maximizar el rendimiento del servidor web TLS. [8]

En otros casos se evalúa el impacto en el rendimiento del uso de TLS como protocolo de transporte para servidores SIP. También evalúa el costo de TLS experimentalmente utilizando un banco de pruebas con OpenSIPS, OpenSSL y Linux ejecutándose en un servidor que está basado en Intel. Se evalúan los costos de TLS como el cifrado masivo de datos, el cifrado de clave pública, el descifrado de clave privada y la verificación basada en MAC. Como resultado se obtuvo que el uso de TLS puede reducir el rendimiento hasta en un factor de 20 en comparación con el caso típico de SIP sobre UDP [9].

Otros trabajos intentan explicar y proponer la forma de evitar o reducir las operaciones criptográficas utilizadas en los mecanismos basados en claves públicas estándar en TLS, bajo un análisis sistemático y una comparación del rendimiento entre los mecanismos de intercambio de claves pre-compartidas y los mecanismos de intercambio de claves públicas. Las métricas de rendimiento fueron el tiempo de procesamiento y la cantidad de datos transmitida para una sesión. Además, se evaluó la interacción entre la duración

global del handshake de TLS y el entorno de red; llegando a la conclusión que al comparar RSA con DHE PSK, este último funciona mejor que RSA solo cuando se usan tamaños de clave pequeños y tienen un rendimiento de red bajo. Aunque DHE PSK puede tener un rendimiento peor que RSA al usar tamaños de clave grandes o alto rendimiento de red, DHE PSK proporciona Perfect Forward Secrecy (PFS) para garantizar una comunicación más segura entre los cifradores de clave pre-compartida [10].

Para el tratamiento de la sobrecarga de solicitudes simultaneas se propuso el balanceo de handshake con el algoritmo EAMRSA que mejora el rendimiento mediante la técnica de transferencia de carga en el protocolo de enlace SSL/TLS. Esta técnica facilita en la distribución de carga favorable al requerir que los clientes realicen más trabajo (como parte de la encriptación) y servidores para realizar un trabajo proporcionalmente menor, lo que resulta en un mejor rendimiento de SSL. Como resultado se obtuvo que el método puede acelerar el procesamiento de las operaciones de claves privada RSA por un factor de entre 4.5 a 18 dependiendo del tamaño de clave RSA [11].

A. Cifrado Web mediante el Protocolo SSL/TLS

El protocolo SSL/TLS ha ido evolucionando desde su implementación en el año 1994 y la publicación de TLSv1.0 en 1999; estos protocolos operan entre la capa de transporte y la de aplicación según el modelo TCP/IP. Las diversas versiones de SSL y TLS ofrecen servicios de seguridad, como la confidencialidad, autenticación de servidor y la integridad del mensaje. Asimismo, estos protocolos se han diseñado considerando características como la eficiencia y extensibilidad, lo que implica un mejor uso de recursos en las comunicaciones reutilizando parámetros de conexión en sesiones simultaneas y la posibilidad de agregar nuevas combinaciones de algoritmos de cifrado además de las predefinidas [12] [13] [14].

B. Rendimiento de un sitio web

Uno de los puntos clave del éxito de un sitio web será el nivel de comodidad de nuestros usuarios, que la experiencia al visitar nuestro sitio sea agradable, que la respuesta que obtengan a sus acciones sea fluida, sin retrasos en las respuestas, etc., esto nos hace afirmar que la percepción del usuario está en función del rendimiento del sitio web en términos de tiempo de procesamiento o respuesta, concurrencia de usuarios, y la seguridad de la comunicación [15] [16] [17] [18].

Los sitios web siguen una arquitectura cliente servidor, donde el cliente es una máquina que solicita un determinado servicio al servidor que es la máquina que lo proporciona. El rendimiento de un sitio web va a depender de ambas partes, ya que el servidor será quien realice el procesamiento de las diversas peticiones haciendo que el consumo de recursos computacionales aumente y por ello el rendimiento del sitio web se vea afectado, así como también depende del cliente que es quien por medio de un navegador consume el servicio web. Este intercambio de mensajes se realiza antes de iniciar con el proceso de transmisión de información en cuestión de segundos, lo que provoca que el tiempo sea superior que cuando la información no fuera cifrada. En ese sentido una comunicación no cifrada además de ser vulnerable va a ser mucho más rápido en tiempo de respuesta. Con la investigación se conoce la diferencia de tiempo y en cuanto puede influir en el rendimiento del sitio web.

La cantidad total de clientes que pueden ser atendidos en simultaneo [17] por un Servidor Web está dada por la formula (1)

$$Max_Clientes = Total_RAM / Max_Proceso_Hijo \quad (1)$$

Asimismo, existen diversas herramientas en línea de fabricantes como Google y Mozilla que permiten medir el rendimiento de un sitio web alojado en internet cada quien, con sus respectivos estándares y valoraciones, en la gran mayoría de los casos nos muestran el tiempo de carga, almacenamiento en cache, peso de la página, tiempo de respuesta, entre otros. Esta información debe de ser considerada ya que evidencia el rendimiento que pueda tener el sitio web.

Existen herramientas como OpenSSL, CypherScan, Wireshark por mencionar los más comunes que permiten obtener información más específica respecto a un sitio web como son: los algoritmos criptográficos utilizados, versiones de TLS aceptados, tiempo de respuesta, peso del sitio web, versión de http utilizado, puertos por defecto, entre otra información.

Asimismo, una forma de medir el consumo de recursos computacionales en función del tiempo de carga. Para calcular el porcentaje de incremento del tiempo de carga de cada sitio web cuando es accedido por HTTPS y HTTP, se empleó la siguiente formula [19].

$$Tasa\ de\ crecimiento\ (\%) = \frac{|Valor\ 1 - Valor\ 2|}{|Valor\ 2|} \times 100\%$$

Valor 1: Población al final del periodo.

Valor 2: Población al principio del periodo.

En esta investigación se reemplazó el Valor 1 por el “Tiempo de carga total de HTTPS” y para valor 2 el “Tiempo de carga total por HTTP” para el mismo sitio web, el cual nos permitió calcular el incremento porcentual existente entre estos dos tiempos.

$$\frac{|Tiempo\ carga\ HTTPS - Tiempo\ carga\ HTTP|}{|Tiempo\ carga\ por\ HTTP|} \times 100\%$$

Como antecedente, dicha fórmula también fue empleada con el objetivo de ver la diferencia de tamaños de datos cifrados y no cifrados que existe cuando el sitio web es accedido por HTTP y HTTPS [20].

Para configurar la versión más adecuada del protocolo SSL/TLS es posible realizarlo dentro del archivo de configuración ssl.conf con la directiva SSLProtocol [21], según la Fig. 5 podemos ver que se han deshabilitado las versiones de SSLv2 y SSLv3, para poder trabajar con todas las versiones de TLS.

Por otro lado, tanto cliente como servidor deben establecer los métodos y algoritmos de cifrado. Esto se realiza durante el proceso de establecimiento de sesión, que es el intercambio de información entre el cliente y servidor que permite lograr un acuerdo de algoritmos y claves a utilizar de manera segura, con la finalidad de garantizar la confidencialidad e integridad. El proceso de negociación básico publicado en el RFC 5246 se muestra en la Fig. 2 [22] [23] [24]

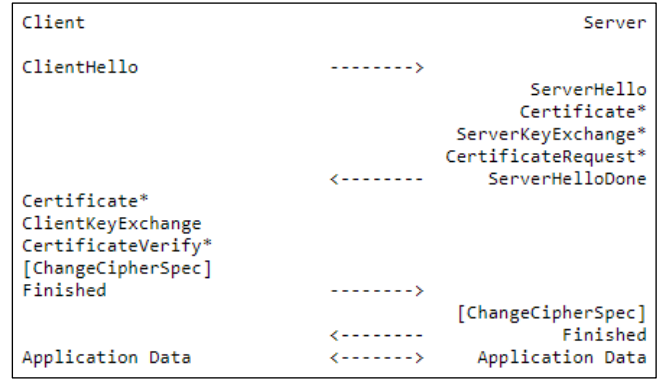


Fig. 3. Proceso de negociación de una sesión SSL/TLS.

Las Autoridades Certificadoras (AC o CA por sus siglas en ingles Certification Authority) son las encargadas de emitir certificados digitales (documento electrónico identificado por un único número de serie con periodo de validez incluido en el propio certificado) el cual permite identificar a determinados equipos y sitios web ante terceros [25].

C. Fortaleza de los algoritmos criptográficos

La fortaleza del algoritmo de cifrado tanto simétrico como asimétrico forman parte de una comunicación por medio de HTTPS. La seguridad y la fuerza del algoritmo de cifrado está relacionada con el tiempo, cada vez más las computadoras son mucho más rápidas y accesibles a un menor costo, permitiendo que los algoritmos de cifrado de mayor complejidad puedan ser procesados con mucha facilidad en comunicaciones por HTTPS. La fortaleza está vinculada con el tamaño de la clave utilizada y por los recursos computacionales con el cual puede ser vulnerado [26].

IV. EVALUACIÓN DEL RENDIMIENTO DE UN SITIO WEB

En esta sección se describirá las pruebas realizadas para evaluar el impacto del uso del protocolo TLS v1.2 en los cinco sitios web que actualmente se encuentran en producción, con el uso de herramientas como Apache Bench, SSLRobot, Qualys SSL Labs.

A. Evaluación con Apache Bench

En la Tabla V se muestra el análisis de rendimiento de los cinco sitios web de producción con Apache Bench (ab), donde se puede observar un mayor tiempo de carga y un menor número de solicitudes atendidas cuando se hace uso del protocolo TLS v1.2 a diferencia del sitio web sin el uso del protocolo de este. El comando utilizado con Apache Bench fue:

#ab -n 1000 -c 25 https://www.web-out.com...(Con SSL/TLS)

ab -n 1000 -c 25 http://www.web-out.com...(Sin SSL/TLS)

Donde:

ab: Comando a usar para el análisis del rendimiento.

-n: Indica el número de solicitudes al servidor.

-c: Indica la cantidad de peticiones concurrentes.

Se realizaron cinco pruebas por cada sitio web con el objetivo de tener un valor promedio con mayor precisión, aumentando la cantidad de peticiones concurrentes para poder evaluar el impacto entre una comunicación cifrada y no cifrada por el protocolo TLS v1.2 tal como se muestra en la Tabla V.

Se consideran algunas características para la prueba, las cuales son:

- Time taken: Tiempo promedio que se mide desde el momento que se crea la primera conexión de socket hasta recibir la última respuesta.
- Requests per Second: Número promedio de solicitudes por segundo, resultado de número de solicitudes entre en tiempo total.

En la Tabla VI Se puede observar el resumen de las pruebas realizadas a los cinco sitios web obteniendo un Time Taken superior y Requests per Second inferior cuando es accedido por HTTPS.

TABLA V. ANÁLISIS DE RENDIMIENTO DEL SITIO WEB-OUT CON APACHE BENCH

WEB1 (WWW.WEB-OUT.COM) - 1,100 Y 500 SOLICITUDES							
SOLICITUDES	# Peticiones Concurrentes	HTTPS		HTTP		Diferencia del Tiempo de Prueba entre HTTP y HTTPS (segundos)	Diferencia de Peticiones por segundo entre HTTP y HTTPS (#/s.)
		Tiempo de la Prueba (segundos)	Media de Peticiones atendidas por segundo (#/s.)	Tiempo de la Prueba (segundos)	Media de Peticiones atendidas por segundo (#/s.)		
1	1	0.278	4.028	0.250	4.024	0.03	0.00
100	20	4.153	24.214	3.801	26.362	0.35	2.15
	40	5.047	20.948	4.067	24.848	0.98	3.90
	60	4.154	24.428	3.907	26.204	0.25	1.78
	80	5.858	18.672	4.352	24.470	1.51	5.80
	100	3.829	26.150	3.659	27.384	0.17	1.23
500	20	24.339	21.070	22.734	22.146	1.61	1.08
PROMEDIO:						0.7	2.277

TABLA VI. TIME TAKEN Y REQUEST PER SECOND ATENDIDAS POR EL SERVIDOR ENTRE HTTP Y HTTPS

Sitio Web	Diferencia de Time Taken de HTTP y HTTPS	% de diferencia de Time Taken de HTTP y HTTPS	Diferencia de Requests per Second per Second HTTP y HTTPS	% de diferencia de Requests per Second per Second HTTP y HTTPS
Web-Out	0.7 s	11%	2.227 #/s	11%
Facultad de Ciencias Económicas y Administrativas de la UNAS	0.25 s	18%	17.82 #/s	23%
Hotel Oro Verde	1.41 s	4%	3.97 #/s	4%
Hotel Natural Green	0.52 s	17%	0.24 #/s	20%
Cámara de Comercio Canadá - Perú	1.08 s	4%	0.12 #/s	4%

Existen diversas combinaciones de valores y técnicas con el cual se pueden obtener diversos resultados para "ab". Para ver más detalles del protocolo SSL/TLS de un sitio web como Cipher Suite y versión de TLS tal como se muestra en Fig. 4,

se utilizó el comando:

```
# ab -n 1 -v 2 https://www.web-out.com
```

```
Transport Protocol      :TLSv1.2
Cipher Suite Protocol  :TLSv1.2
Cipher Suite Name      :ECDHE-RSA-AES256-GCM-SHA384
Cipher Suite Cipher Bits:256 (256)
SSL-Session:
Protocol               : TLSv1.2
Cipher                 : ECDHE-RSA-AES256-GCM-SHA384
Session-ID             : 1499055C8EF11117F18338CC22BCE5C30DAFF8C31E6F2977074019BC77713DA2
```

Fig. 4. Información de Protocolo, Suite de Cifrado, ID-Session y valides de certificado del sitio web https://www.web-out.com.

Es importante conocer cuál es la suite de cifrado, validez del certificado digital, la versión de protocolo que está habilitado, entre otros datos. Esto permite al administrador tomar medidas correctivas si es necesario.

Para identificar las suites de cifrado y su prioridad al momento de ingresar al sitio web de web-out, se usó la herramienta Cipherscan desarrollada por Mozilla tal como muestra la Fig. 5.

```
Target: www.web-out.com:443

prio  ciphersuite                protocols                pfs
1      ECDHE-RSA-AES256-GCM-SHA384 TLSv1.2                  ECDH,8-571,570bits
2      ECDHE-RSA-AES256-SHA384    TLSv1.2                  ECDH,8-571,570bits
3      ECDHE-RSA-AES256-SHA       TLSv1,TLSv1.1,TLSv1.2  ECDH,8-571,570bits
4      DHE-RSA-AES256-GCM-SHA384  TLSv1.2                  DH,2048bits
5      DHE-RSA-AES256-SHA256     TLSv1.2                  DH,2048bits
```

Fig. 5. Prioridad de Suite de Cifrado SSL/TLS del sitio web https://www.web-out.com

Para automatizar el trabajo de obtención de la información con la herramienta Apache Bench (ab) de ambos sitios web, se creó un script con bash, cuyo código es el siguiente:

```
#!/bin/bash
echo "Análisis de la Información de Test de Rendimiento"
echo "-----"
echo "Quiere realizar un analisis del sitio:"
echo "1 -> Análisis con TLS v1.2"
echo "2 -> Análisis sin TLS v1.2"
echo "-----"
read var1
if [ $var1 -eq "1" ]; then
    echo "Análisis con TLSv1.2"
    echo "-----"
    echo "Ingrese el Sitio Web a evaluar:"
    echo "1 - https://www.web-out.com/"
    echo "2 - https://www.fceaunas.edu.pe/"
    echo "3 - https://www.canadaperu.org/"
    echo "4 - https://www.hotel-oroverde.com/"
    echo "5 - https://www.hotelnaturalgreen.com/"
    echo "-----"
    read site
    echo "-----"
    case $site in
        1) siteeval="https://www.web-out.com/"
        ;;
        2) siteeval="https://www.fceaunas.edu.pe/"
        ;;
        3) siteeval="https://www.canadaperu.org/"
        ;;
        4) siteeval="https://www.hotel-oroverde.com/"
        ;;
        5) siteeval="https://www.hotelnaturalgreen.com/"
        ;;
        *)
        esac
    echo "El sitio a evaluar es: '$siteeval'"
    rm -f resumen_con_ssl.txt
    array=(20 40 60 80 100)
    echo "Iniciando Test..."
    for i in ${array[@]};
    do
        ab -n 100 -c $i $siteeval > web_con_ssl_${i}.txt;
    done
    echo $i
```

```

done
for i in $(array[@]);
do
ls | grep "Concurrency Level" web_con_ssl_$i.txt >> resumen_con_ssl.txt
ls | grep "Time taken for tests" web_con_ssl_$i.txt >> resumen_con_ssl.txt
ls | grep "Requests per second" web_con_ssl_$i.txt >> resumen_con_ssl.txt
echo "-----" >> resumen_con_ssl.txt
done
fi

if [ $var1 -eq "2" ]; then
echo "Análisis sin TLS v1.2"
echo "-----"
echo " Ingrese el Sitio Web a evaluar:"
echo "1 - http://www.web-out.com/"
echo "2 - http://www.fceaunas.edu.pe/"
echo "3 - http://www.canadaperu.org/"
echo "4 - http://www.hotel-oroverde.com/"
echo "5 - http://www.hotelnaturalgreen.com/"
echo "-----"
read site
echo "-----"
case $site in
1) siteeval="http://www.web-out.com/"
;;
2) siteeval="http://www.fceaunas.edu.pe/"
;;
3) siteeval="http://www.canadaperu.org/"
;;
4) siteeval="http://www.hotel-oroverde.com/"
;;
5) siteeval="http://www.hotelnaturalgreen.com/"
;;
esac
echo "El sitio a evaluar es:" $siteeval
rm -f resumen_sin_ssl.txt
array=(20 40 60 80 100)
echo "Iniciando Test..."
for i in $(array[@]);
do
ab -n 100 -c $i $siteeval > web_sin_ssl_$i.txt;
done
done
for i in $(array[@]);
do
ls | grep "Concurrency Level" web_sin_ssl_$i.txt >> resumen_sin_ssl.txt
ls | grep "Time taken for tests" web_sin_ssl_$i.txt >> resumen_sin_ssl.txt
ls | grep "Requests per second" web_sin_ssl_$i.txt >> resumen_sin_ssl.txt
echo "-----" >> resumen_sin_ssl.txt
done
fi

```

B. Tiempo de procesamiento

El uso del Protocolo SSL/TLS genera un tiempo de latencia el cual genera un tiempo superior cuando un sitio web está siendo accedido por HTTPS, esto debido al proceso de negociación que existe entre el cliente y servidor, en la Fig. 6 se describe tiempos aproximados en ms para cada intercambio de información entre cliente y servidor.

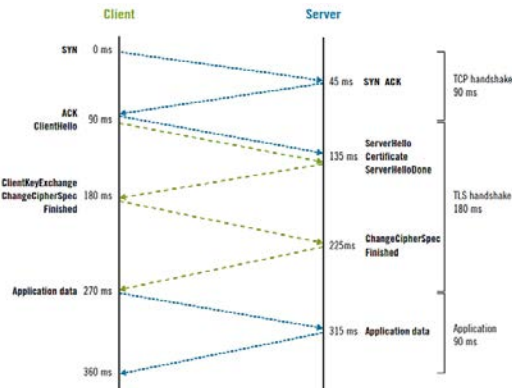


Fig. 6. Latencias de TCP Handshake y TLS Handshake

El tiempo de carga de los cinco sitios web cuando es accedido por HTTP y HTTPS se detallan en la Tabla VII. En la que se puede apreciar un leve incremento cuando el sitio web es accedido por HTTPS, esta información ha sido obtenida mediante la herramienta de desarrollo del navegador Mozilla Firefox.

TABLA VII. DIFERENCIA DE NÚMERO DE SOLICITUDES Y TIEMPO DE CARGA TOTAL DEL SITIO WEB.

Sitio Web	Núm. de Solicitud es HTTP	Núm. de Solicitud es HTTPS	Tiempo en cargar todas las solicitudes HTTP	Tiempo en cargar todas las solicitudes HTTPS	% de Incremento del Tiempo de cada de los sitios web
web-out.com	86	86	8.928 s	10.300 s	15%
fceaunas.edu.pe	106	106	6.386 s	6.500 s	2%
hotel-oroverde.com	94	94	18.692 s	19.842 s	6%
hotelnaturalgreen.com	107	107	9.572 s	9.906 s	3%
canadaperu.org	146	146	20.234 s	21.324 s	5%

C. Algoritmos de cifrado

Una de las consideraciones importantes para el uso adecuado del protocolo TLS v1.2 es la selección de los algoritmos de cifrado robustos y descartar aquellos que no presentan tal característica. Algunas herramientas como “Analyze” (desarrollada en lenguaje Python y disponible como parte de la suite de Cipherscan) ayudan a identificar los algoritmos que se deberían descartar. El resultado de ejecutar el comando “./analyze.py -t https://www.web-out.com” con esta herramienta en el servidor web se muestra en la Fig. 7.

```

Changes needed to match the old level:
* enable SSLv3
* use a certificate with sha1withRSAEncryption signature
* use DHE of 1024bits and ECC of 160bits
* enforce server side ordering

Changes needed to match the intermediate level:
* consider using DHE of at least 2048bits and ECC 256bit and greater
* enforce server side ordering
* increase priority of DHE-RSA-AES256-GCM-SHA384 over ECDHE-RSA-AES256-SHA
* increase priority of ECDHE-RSA-AES128-GCM-SHA256 over AES256-SHA
* increase priority of DHE-RSA-AES128-GCM-SHA256 over ECDHE-RSA-AES128-SHA
* increase priority of ECDHE-RSA-DES-CBC3-SHA over AES128-SHA
* fix ciphersuite ordering, use recommended intermediate ciphersuite

Changes needed to match the modern level:
* remove cipher ECDHE-RSA-AES256-SHA
* remove cipher DHE-RSA-AES256-GCM-SHA384
* remove cipher DHE-RSA-AES256-SHA256
* remove cipher DHE-RSA-AES256-SHA
* remove cipher AES256-GCM-SHA384
* remove cipher AES256-SHA256
* remove cipher AES256-SHA
* remove cipher ECDHE-RSA-AES128-SHA
* remove cipher DHE-RSA-AES128-GCM-SHA256
* remove cipher DHE-RSA-AES128-SHA256
* remove cipher DHE-RSA-AES128-SHA
* remove cipher AES128-GCM-SHA256
* remove cipher AES128-SHA256
* remove cipher AES128-SHA
* remove cipher ECDHE-RSA-DES-CBC3-SHA
* remove cipher EDH-RSA-DES-CBC3-SHA
* remove cipher DES-CBC3-SHA
* disable TLSv1.1
* disable TLSv1

```

Fig. 7. Recomendaciones de suites de cifrado a modificar en el Servidor Web.

Asimismo, es posible considerar las recomendaciones de Mozilla, en la que se puede obtener la configuración moderna, intermedia y antigua relacionadas con las versiones de software de servidor y de OpenSSL.

En la Tabla VIII se detalla información del certificado digital obtenido con la herramienta SSL Robot de los cinco sitios web, indicando el tipo y tamaño de clave pública, la Autoridad Certificadora y el algoritmo de firma del Certificado Digital generado para cada sitio web.

Se puede apreciar una similitud en los cinco sitios web que emplean el algoritmo RSA con un tamaño de clave de 2048 bits. Para el campo de algoritmo de firma todos los sitios web emplean como algoritmo de hash a SHA-256 excepto para el sitio web del Hotel Oro Verde que emplea SHA-384, y como algoritmo de firma es RSA-2048.

TABLA VIII. CERTIFICADO DIGITAL EMPLEADO EN LA CONEXIÓN CON LOS CINCO SITIOS WEB POR HTTPS

	Clave	CA	Algoritmo de Firma
Web-Out	RSA 2048	COMODO RSA Domain Validation Secure Server CA	SHA 256 WITH RSA
Facultad de Ciencias Económicas y Administrativas	RSA 2048	COMODO RSA Domain Validation Secure Server CA	SHA 256 WITH RSA
Hotel Oro Verde	RSA 2048	CPANEL INC. Certification Authority	SHA 384 WITH RSA
Hotel Natural Green	RSA 2048	COMODO RSA Domain Validation Secure Server CA	SHA 256 WITH RSA
Cámara de comercio Canadá - Perú	RSA 2048	COMODO RSA Domain Validation Secure Server CA	SHA 256 WITH RSA

D. Evaluación con Qualys SSL Labs

Es una herramienta completa que muestra información respecto al nivel de seguridad que cuenta un sitio web respecto a su configuración del protocolo SSL/TLS, funciona online y asigna una valoración del sitio web respecto a tres criterios: “Soporte de Protocolo”, “Intercambio de llaves” y “Fuerza de cifrado”.

Categoría	Puntuación	Numerical Score	Grade
Soporte de protocolo	30%	score >= 80	A
Intercambio de llaves	30%	score >= 65	B
Fuerza de cifrado	40%	score >= 50	C
		score >= 35	D
		score >= 20	E
		score < 20	F

Fig. 8. Puntuación a las categorías para SSL Labs.

En la Tabla IX se puede visualizar la calificación final obtenida por la herramienta Qualys SSL Labs para los cinco sitios web, existiendo un intervalo de calificación de A-F, siendo A la nota más alta y F la nota más baja respecto a los criterios descritos en la Fig. 9.

TABLA IX. CALIFICACIÓN OBTENIDA POR QUALYS SSL LABS PARA LOS CINCO SITIOS WEB

Sitio Web	Calificación General
Web-Out	B
Facultad de Ciencias Económicas y Administrativas - UNAS	B
Hotel Oro Verde	A
Hotel Natural Green	B
Cámara de Comercio Canadá - Perú	B

E. Evaluación con OpenSSL

En la Tabla X se puede visualizar que el algoritmo RSA conforme aumenta la longitud de clave, el proceso de firmado es mucho más lento que el proceso de verificación de la firma. Así mismo nos indica que el servidor mediante el algoritmo

RSA 2048 bits es posible realizar 331.8 firmas/segundo. Con el algoritmo ECDSA 256 bits es posible 3572 conexiones TLS por segundo con un proceso de verificación mucho más lento que usando RSA.

TABLA X. RENDIMIENTO DEL ALGORITMO DE FIRMA RSA, DSA Y ECDSA EN UNA COMUNICACIÓN HTTPS

Algoritmo/Longitud de clave	Firma	Verificación	Firmas /s.	Verificación/s.
RSA 512 bits	0.000625 s.	0.000030 s.	1599.8	33411.8
RSA 1024 bits	0.000812 s.	0.000069 s.	1231.4	14534.3
RSA 2048 bits	0.003014 s.	0.000203 s.	331.8	4922.8
RSA 3072 bits	0.021355 s.	0.000439 s.	46.8	2277.7
RSA 4096 bits	0.042895 s.	0.000699 s.	23.3	1430.7
RSA 7680 bits	0.364444 s.	0.002103 s.	2.7	475.6
DSA 512 bits	0.000997 s.	0.000572 s.	1003.1	1749
DSA 1024 bits	0.001378 s.	0.001071 s.	725.5	934.1
DSA 2048 bits	0.0081 s.	0.0056 s.	202	433.7
ECDSA 256 bits (nistp256)	0.0003 s.	0.0007 s.	3572.1	1512
ECDSA 384 bits (nistp384)	0.0189 s.	0.0138 s.	52.8	72.5

V. RESULTADOS

De acuerdo con los resultados presentados en la Tabla VI sobre las pruebas con Apache Bench, los valores de tiempo de carga (time taken) el protocolo SSL/TLS no influye significativamente en el nivel de solicitudes que pueda atender un servidor web, habiendo realizado 5 pruebas de saturación del servidor para cada sitio web, existiendo un total de 25 pruebas mediante la herramienta ApacheBench (ab) donde se realizaron 1, 100 y 500 solicitudes con una concurrencia de peticiones de 1, 20, 40, 60, 80 y 100 hacia el VPS 1 y VPS 2. La variación de los tiempos de carga y la velocidad de respuesta por solicitud no varían significativamente existiendo un margen de diferencia de 1.5 segundos adicionales para las conexiones por HTTPS (TLS versión 1.2) y existiendo una mayor velocidad de atención de solicitudes por medio de HTTP, siendo mucho más rápido en un 11% para el sitio web de Web-Out S.A.; un 23% para el sitio web de la Facultad de Ciencias Económicas y Administrativas de la UNAS; un 4% para el sitio web del Hotel Oro Verde; un 20% para el sitio web del Hotel Natural Green y un 4% para la cámara de Comercio de Canadá Perú. Además del impacto del protocolo TLS versión 1.2 en el rendimiento de los sitios web.

Asimismo, el resultado de las pruebas de nivel de seguridad del sitio web con la herramienta Qualys SSL Labs indican que los cinco sitios web obtienen una puntuación de entre A y B lo cual se consideran configuraciones optimas, teniendo un impacto positivo en la seguridad de la información transmitida durante la interacción del sitio web con el usuario final.

Sin embargo, se debe considerar que desde la perspectiva del usuario promedio es posible que no se perciba las diferencias encontradas como poco significativas debido a la potencia de hardware actualmente disponible.

VI. CONCLUSIONES

El uso del protocolo TLS versión 1.2 para el cifrado de sitios web (HTTPS) es altamente recomendable, debido a los riesgos que representa el transferir información en texto plano con protocolos no cifrados como HTTP. Aunque el uso de protocolos como TLS generan mayor consumo de recursos computacionales afectando mínimamente el rendimiento de los sitios web, expresados en tiempo de respuesta o carga (time taken), la percepción negativa del usuario promedio es posible que no sea significativa.

Se ha podido comprobar que mientras más componentes (imágenes, archivos css, archivos java script, videos, entre otros) tenga un sitio web, el número de solicitudes HTTP o HTTPS se incrementaran, reflejándose en un mayor tiempo de carga. Se recomienda reducir en lo más mínimo estos componentes ya sea combinando archivos y script en archivos únicos, o eliminando y reduciendo el peso de componentes innecesarios, así mismo empleando el guardado en cache.

Asimismo, se puede demostrar que el uso del protocolo SSL/TLS influye de manera positiva en la seguridad de los sitios web y esto se puede comprobar mediante la herramienta Qualys SSL Labs obteniendo una puntuación (A-F) respecto a las configuraciones habilitadas en el servidor web, se debe considerar que si la configuración de este protocolo es incorrecta será igual de vulnerable como si la navegación fuera por HTTP.

Además del impacto del protocolo TLS versión 1.2 en el rendimiento de los sitios web, es importante considerar los tipos de algoritmos de cifrado y la combinación de uso para garantizar la mejor protección durante una sesión e intercambio de información a nivel de la web.

TRABAJOS FUTUROS

Realizar evaluaciones incluyendo otros servidores Web como IIS de Microsoft y GlasFish, además de considerar sitios web mas complejos que realicen transacciones con bases de datos.

Finalmente, complementar la evaluación de este tipo de protocolos en los sitios web, poniendo a prueba su fortaleza y robustez con adecuadas técnicas de pruebas de penetración de forma controlada.

Comparar el rendimiento de sitios web entre el uso del protocolo TLSv1.2 y TLSv1.3.

AGRADECIMIENTOS

A la empresa Web-Out por permitir analizar cinco de sus sitios web desarrollados y administrados.

REFERENCIAS

- [1] J. D. Irwin y C.-H. Wu, «Introduction to Computer Networks and Cybersecurity,» 2013.
- [2] IETF, «Who we are,» [En línea]. Available: <https://www.ietf.org/about/who/>. [Último acceso: 15 Noviembre 2018].
- [3] E. Rescorla, «The Transport Layer Security (TLS) Protocol,» Agosto 2008. [En línea]. Available: <https://www.ietf.org/rfc/rfc5246.txt>.

- [4] E. Rescorla, «HTTP Over TLS,» Network Working Group, 2000. [En línea]. Available: <https://tools.ietf.org/html/rfc2818>.
- [5] D. Holmes, «The 2017 TLS Telemetry Report,» 23 Abril 2018. [En línea]. Available: <https://www.f5.com/labs/articles/threat-intelligence/the-2017-tls-telemetry-report>.
- [6] Qualys Inc., «SSL Labs,» 09 Abril 2019. [En línea]. Available: <https://www.ssllabs.com/ssl-pulse/>. [Último acceso: abril 2019].
- [7] OpenSSL, «TLS/SSL and crypto library,» [En línea]. Available: <https://github.com/openssl/openssl>. [Último acceso: Octubre 2018].
- [8] C. Coarfa, P. Druschel y D. Wallach, «Performance Analysis of TLS Web Servers,» *IEEE Xplore Digital Library*, pp. 39-69, 2006.
- [9] C. Shen, E. Nahum, H. Schulzrinne y C. Wright, «The Impact of TLS on SIP Server Performance,» 2009.
- [10] F.-C. Kuo, H. Tschofenig y F. Meyer, «Comparison Studies between Pre-Shared and Public Key Exchange Mechanisms for Transport Layer Security,» *IEEE Xplore Digital Library*, 2006.
- [11] H. Li y G. Zhao, «Improving Secure Server Performance By EAMRSA SSLHandshakes,» *IEEE Xplore Digital Library*, 2012.
- [12] IETF, *The TLS Protocol version 1.0*, vol. RFC 2246, IETF, 1999.
- [13] R. Oppliger, *SSL and TLS Theory and Practice*, London: Artech House, 2016.
- [14] M. López Fernández, «Caracterización y medida pasiva del rendimiento para conexiones Web seguras HTTPS,» España, 2015.
- [15] C. Mateu, *Desarrollo de Aplicaciones Web*, Barcelona, 2004.
- [16] F. Carvajal Palomares, «Administración y Auditoría de los servicios web,» Editorial CEP, S.L., 2017.
- [17] J. L. Villada Romero, *Instalación y configuración del software de servidor web (UF1271)*, IC Editorial, 2015, p. 403.
- [18] J. Sabogal Rosas, «Modelamiento de una plataforma virtual para la gestión de avisos normativos y de trámite legal,» 2015.
- [19] A. O. Pallmall, *Demografía, un problema global*, 2014.
- [20] R. A. Ariansen Moncada y J. I. Rojas Diaz, «Implementación de protocolo de cifrado TLS para mejorar la seguridad de las comunicaciones en la capa de transporte 2016,» Chiclayo, 2016.
- [21] Apache, «Apache Modulo mod_ssl,» [En línea]. Available: https://httpd.apache.org/docs/2.4/mod/mod_ssl.html#sslprotocol.
- [22] IETF, «The Transport Layer Security (TLS) Protocol version 1.2,» agosto 2008. [En línea]. Available: <https://tools.ietf.org/html/rfc5246>.
- [23] M. Driscoll, «The Illustrated TLS Connection v1.2,» diciembre 2018. [En línea]. Available: <https://tls.ulfheim.net/>.
- [24] M. Driscoll, «The New Illustrated TLS Connection v1.3,» diciembre 2018. [En línea]. Available: <https://tls13.ulfheim.net/>.
- [25] G. Escrivá Gascó, R. Romero Serrano, D. Jorge Ramada y R. Onrubia Pérez, *Seguridad Informática*, Madrid, 2013.
- [26] I. Ristic, *Bulletproof SSL and TLS: The Complete Guide to Deploying Secure Servers and Web Applications*, 2014.

William-Rogelio Marchand-Niño, Ingeniero de Sistemas otorgado por la Universidad Nacional del Centro del Perú, con maestría en Dirección Estratégica de TI de la Universidad de Piura, con 18 años de experiencia académica en UNAS, UDH, UPLA. Desde el año 2004 es profesor asociado en la UNAS. Ha impartido más de 90 cursos de pregrado en diferentes universidades. Instructor CISCO por 12 años. Posee múltiples certificaciones de la Industria como PMP, ITIL Foundation, CCNA, MTA. Director del Centro de Tecnologías de Información y Comunicación de la Universidad Nacional Agraria de la Selva. Miembro Senior de la IEEE.

Edgar Etson Rueda Liberato, Bachiller en Ingeniería en Informática y Sistemas de la Universidad Nacional Agraria de la Selva (UNAS). Miembro del Grupo de Investigación de Redes y Seguridad de la Facultad de Ingeniería en Informática y Sistemas (FIIS-UNAS)

TENDENCIAS DE LA EDUCACIÓN EN SEGURIDAD

Laboratory for Vulnerability Analysis and CIS Controls on Layer 2 Switches

William-Rogelio Marchand-Niño¹
william.marchand@unas.edu.pe

José Martín Santillán Ruiz¹
jose.santillan@unas.edu.pe

Resumen— En el proceso de enseñanza de temas relacionados con la seguridad de la red, es importante que el estudiante tenga un enfoque más real de las vulnerabilidades y los tipos de ataques que pueden especificarse en los entornos de red de producción de infraestructuras de Tecnologías de Información y Comunicación. El objetivo principal es proporcionar al alumno escenarios que le permitan llevar a cabo un análisis de vulnerabilidad completo, desde la identificación de la vulnerabilidad, explotación o verificación de la vulnerabilidad con técnicas de *hacking* ético, evaluación CVSS, mitigación y alineación con algún control de CIS (*Center of Internet Security*). El Laboratorio propuesto es una forma de apoyo para el proceso de enseñanza-aprendizaje de la seguridad de la red a nivel universitario. Ofrece un paquete práctico que incluye las fases generales y comunes de un ataque (enfoque de seguridad ofensiva), procedimientos de mitigación (enfoque de seguridad defensiva) y controles de seguridad basados en una referencia válida como CIS (gestión de seguridad informática).

Palabras clave— Vulnerabilidades, *pentesting*, redes, CIS, CVSS, capa dos OSI, amenazas, seguridad ofensiva, aprendizaje.

I. INTRODUCCIÓN

La investigación está orientada a la propuesta de un Laboratorio para el análisis de las vulnerabilidades asociadas a la deficiente configuración de los dispositivos de red de capa 2, específicamente los conmutadores (switches) de red. La deficiente configuración o las omisiones en las misma, también se debe al efecto de crecimiento de la red, al incrementarse el número de dispositivos, también se incrementa el volumen de configuraciones, y esto genera que las amenazas (de reconocimiento, acceso o denegación de servicio) puedan concretarse en ataques o incidentes de seguridad no deseados dañando a los sistemas u organización. [1] [2] [3].

Para la definición del alcance del Laboratorio se identificó las principales vulnerabilidades que presentan comúnmente los dispositivos de capa de enlace de datos, tomando como referencia la base de datos de dominio público CVE (Common Vulnerabilities and Exposures) siendo algunas de estas, ARP spoofing, salto de VLAN, ataques de fuerza bruta o diccionario contra servicios TELNET y SSH, ataques de hombre-en-el-medio, DHCP spoofing y claim role STP forman parte de las pruebas de concepto con los recursos disponibles para el montaje de un Laboratorio orientado al estudio de seguridad de redes [4] [5] [6]. Asimismo, se consideró incluir la valoración del grado de severidad que está de acuerdo con el Sistema de Calificación de Vulnerabilidades Comunes (CVSS) [7]. El objetivo principal es proveer al estudiante escenarios que pueden presentarse en entornos de producción y realizar un análisis de vulnerabilidades

completo, desde la identificación de la vulnerabilidad, explotación o verificación de la vulnerabilidad con técnicas de ethical hacking (enfoque de seguridad ofensiva), valoración CVSS, mitigación (enfoque de seguridad defensiva), y alineación con los 20 controles (enfoque de la gestión de seguridad informática) críticos de CIS (Center for Internet Security) [8].

Para el desarrollo del Laboratorio se ha formulado una rubrica asociada para la verificación del cumplimiento de los pasos formulados.

Una característica para resaltar de la propuesta de este Laboratorio es su forma de aplicación, que está orientada a un tipo de desafío o CtF (Capture the Flag) con formación de equipos, que en otros estudios muestran un impacto positivo, lo que indica en principio, que es un método efectivo para afianzar las capacidades y habilidades en temas relacionados a la seguridad informática, promoviendo el trabajo en equipo, la colaboración y la competencia [9].

La práctica de evaluar o diagnosticar el nivel de seguridad de una red debe ser una práctica común y frecuente, porque a pesar que los sistemas puedan tener instalado las últimas actualizaciones no significa que están libres de vulnerabilidades, por el contrario, se evidencia que las deficientes configuraciones son causa de vulnerabilidades que son aprovechadas por los actores de amenaza [10].

Una de las formas de evaluar la robustez de la infraestructura es sometiendo a prueba las defensas implementadas a nivel de configuraciones y servicios habilitados en los equipos de red. Uno de los beneficios de utilizar pruebas de penetración en el contexto de la seguridad informática es que esta provee de un enfoque desde la perspectiva de un atacante real que dirige técnicas de explotación para romper la protección de un sistema [11] [12].

A nivel de los switches que operan en la capa de enlace de datos de acuerdo al modelo de referencia OSI, se establece algunos de los ataques más frecuentes dirigidos a este tipo de equipos, entre los que destacan, saturación de direcciones MAC, aprovechamiento de protocolos de descubrimiento, suplantación de identidad de switch, ataque de etiquetado doble, y la denegación de servicio [13] [14], que a su vez están relacionados a los siguientes tipos de vulnerabilidades:

- Imperfecciones en las políticas.
- Errores de diseño
- Deficiencias de protocolos
- Deficiencias en la configuración

¹ Grupo de Investigación en Redes, Seguridad y Gestión de TI
Universidad Nacional Agraria de la Selva - Tingo María, Perú

- Debilidades en el software
- Factores humanos
- Software malicioso
- Vulnerabilidades de hardware
- Acceso físico a los recursos de red.
- Cifrado y autenticación

De la relación anterior y para efectos del trabajo de investigación, se consideran las siguientes:

- Deficiencias de protocolos
- Deficiencias en la configuración
- Cifrado y autenticación.

La fase de explotación en el contexto de las pruebas de penetración es la aplicación de exploits para lograr el acceso a los sistemas de un cliente. También es aprovechar las vulnerabilidades identificadas previamente, es la real fase de ataque. No es necesario utilizar exploits (código malicioso) para realizar el proceso de explotación. Otra definición de la explotación es el aprovechamiento de los fallos lógicos de los sistemas informáticos para lograr acceso privilegiado a la red, extraer información sensible o persistencia de ataque, mediante herramientas o técnicas de explotación como ataques de fuerza bruta o diccionario, ejecución de código, ingeniería social, pivoting, etc. [15] [16] [17].

Al realizar la explotación de vulnerabilidades se pueden utilizar herramientas automatizadas, sin embargo, se debe tener conciencia y certeza de lo que realmente está ejecutando el código de la herramienta o exploit. Muchas veces la automatización hace que el pentester pierda cierto control sobre el proceso de explotación, por lo que un pentester debe conocer y tener certeza de lo que se ejecuta y el tipo de vulnerabilidad que se está analizando [18].

En el aspecto de formas de mitigación, las soluciones propuestas en la literatura para enfrentar los ataques y amenazas en redes LAN Ethernet abarcan entre otros, el reemplazo de switch por router; protección física, segmentación VLAN; control de acceso de host basado en autenticación 802.1x; listas de control de acceso; seguridad de puerto; protección de sobrecarga, seguridad centralizada, protocolos seguros; monitorización de seguridad; además de corregir errores cometidos por parte de los administradores en los procedimientos de configuración [19].

La valoración de las vulnerabilidades se realizó utilizando la calculadora de CVSS versión 3 que tiene las siguientes métricas base [7]:

- Vector de Ataque (AV): Red, Adyacente, Local, Físico.
- Complejidad de Ataque (AC): Bajo, Alto.
- Privilegios Requeridos (PR): Ninguno, Bajo, Alto.
- Interacción con Usuario (UI): Ninguno, Requerido.
- Alcance (S): Sin cambios, cambiado.
- Confidencialidad (C): Ninguno, Bajo, Alto
- Integridad (I): Ninguno, Bajo, Alto.

- Disponibilidad (A): Ninguno, Bajo, Alto.

El resultado de la valoración estable una clasificación del nivel de severidad de la vulnerabilidad que se muestra en la Tabla I.

TABLA I. NIVELES DE SEVERIDAD CVSS

Nivel de severidad de vulnerabilidad	Puntuación CVSS
Ninguno	0.0
Bajo	0.1 – 3.9
Medio	4.0 – 6.9
Alto	7.0 – 8.9
Crítico	9.0 – 10.0

II. TIPOS DE ATAQUES PARA VULNERABILIDADES CONTRA SWITCHES

Algunos tipos de ataques o pruebas de concepto asociadas se describen a continuación:

A. Suplantación de identidad de DHCP.

DHCP es el protocolo que asigna automáticamente una dirección IP válida de un pool de DHCP a un host. Se pueden realizar dos tipos de ataques DHCP a una red conmutada: los ataques de agotamiento de DHCP y los de suplantación de identidad de DHCP.

En los ataques de suplantación de identidad de DHCP, un atacante configura un servidor de DHCP falso en la red para asignar direcciones de DHCP para los clientes. El motivo común de este ataque es obligar a los clientes a que usen servidores de Sistema de nombres de dominios (DNS) o de Servicio de nombres Internet de Windows (WINS) falsos y hacer que los clientes usen al atacante, o una máquina controlada por el atacante como gateway predeterminado.

B. Aprovechamiento de CDP

El Protocolo de Descubrimiento de Cisco (CDP, Cisco Discovery Protocol) es un protocolo propiedad de Cisco que puede configurarse en todos los dispositivos de este fabricante. CDP detecta otros dispositivos de Cisco conectados directamente, lo que permite que los dispositivos configuren su conexión de forma automática. En algunos casos, esto simplifica la configuración y la conectividad.

De manera predeterminada, la mayoría de los routers y switches Cisco poseen CDP habilitado en todos los puertos. La información de CDP se envía en broadcast periódicas sin cifrar. Esta información se actualiza localmente en la base de datos de CDP de cada dispositivo. Debido a que CDP es un protocolo de capa 2, los routers no propagan los mensajes CDP.

El protocolo análogo al CDP es el protocolo LLDP (Link Layer Discovery Protocol) que tiene funciones similares de descubrimiento de dispositivos a nivel de capa de enlace de datos. Este protocolo a diferencia de CDP es “multivendor”, es decir opera de forma independiente a la marca del dispositivo.

El aprovechamiento de CDP se asocia a la divulgación información por la operación del protocolo en los puertos en los que no debería estar activo, y desde el punto de vista de los actores de amenazas se denomina un ataque reconocimiento.

Para efectos de las pruebas y por disponibilidad de equipos Cisco se utilizó el protocolo CDP.

C. Ataque de suplantación de identidad de Switch

En un ataque de suplantación de identidad de switch básico, el atacante aprovecha la configuración predeterminada del puerto del switch establecido en dinámico automático. El atacante de la red configura un sistema para suplantar su propia identidad y hacerse pasar por un switch. Esta suplantación de identidad requiere que el atacante de la red pueda emular mensajes 802.1Q y DTP. Al engañar al switch que otro switch intenta crear un enlace troncal, el atacante puede acceder a todas las VLAN permitidas en el puerto de enlace troncal.

Los saltos de VLAN permiten que una VLAN pueda ver el tráfico de otra VLAN. La suplantación de identidad de switch es un tipo de ataque con salto de VLAN que funciona mediante el aprovechamiento de un puerto de enlace troncal mal configurado. De manera predeterminada, los puertos de enlace troncal tienen acceso a todas las VLAN y pasan el tráfico para varias VLAN a través del mismo enlace físico, generalmente entre switches.

D. Ataque de etiquetado doble

Este tipo de ataque aprovecha la forma en que funciona el hardware en la mayoría de los switches. La mayoría de los switches realizan solo un nivel de desencapsulación 802.1Q, lo que permite que un atacante incorpore una etiqueta 802.1Q oculta en la trama. Esta etiqueta permite que la trama se reenvíe a una VLAN que la etiqueta 802.1Q original no especificó. Una característica importante del ataque con salto de VLAN de encapsulado doble es que funciona incluso si se inhabilitan los puertos de enlace troncal, ya que, generalmente, un host envía una trama por un segmento que no es un enlace troncal.

Los ataques con salto de VLAN de etiquetado doble implican los siguientes tres pasos:

- El atacante envía una trama 802.1Q con doble etiqueta al switch. El encabezado externo tiene la etiqueta VLAN del atacante, que es la misma que la VLAN nativa del puerto de enlace troncal. Se supone que el switch procesa la trama que recibe del atacante como si estuviera en un puerto de enlace troncal o un puerto con una VLAN de voz (un switch no debe recibir una trama de Ethernet etiquetada en un puerto de acceso). Por ejemplo, suponga que la VLAN nativa es la VLAN 10. La etiqueta interna es la VLAN víctima; en este caso, la VLAN 20.
- La trama llega al switch, que observa la primera etiqueta 802.1Q de 4 bytes. El switch observa que la trama está destinada a la VLAN 10, que es la VLAN nativa. El switch reenvía el paquete por todos los puertos de la VLAN 10 después de eliminar la etiqueta de VLAN 10. En el puerto de enlace troncal, se elimina la etiqueta de VLAN 10, y no se vuelve a etiquetar el paquete porque esta forma parte de la VLAN nativa. En este punto, la etiqueta de VLAN 20 sigue intacta, y el primer switch no la inspeccionó.
- El segundo switch observa solo la etiqueta 802.1Q interna que envió el atacante y ve que la trama está destinada a la VLAN 20, el objetivo. El segundo switch envía la trama al puerto víctima o lo satura, según si existe una entrada en la tabla de direcciones MAC para el host víctima.

Este tipo de ataque es unidireccional y solo funciona cuando el atacante se conecta a un puerto que reside en la misma VLAN que la VLAN nativa del puerto de enlace troncal.

III. METODOLOGÍA

En esta sección se detallará el proceso de las pruebas de concepto, la valoración y mitigación asociada con los controles de CIS.

Para el montaje del Laboratorio se utilizaron los siguientes equipos:

- 3 switches Cisco Catalyst 2960 con IOS versión 12.
- Servidor de pruebas con sistema operativo Linux Ubuntu 16.04
- Dos máquinas virtuales con sistema operativo Windows 7. (víctimas)
- Una máquina virtual con sistema operativo Kali Linux 2018. (atacante)

El proceso de desarrollo de pruebas de concepto se realizó sobre una topología básica de experimentación con los equipos indicados anteriormente. La topología se muestra en la Fig. 1.

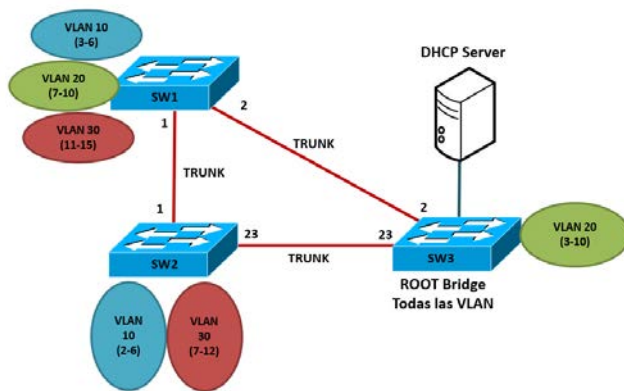


Fig. 1. Topología base para las Pruebas de Concepto.

De acuerdo con las vulnerabilidades comunes para los equipos de la capa de enlace de datos se definieron las pruebas específicas mostradas en la Tabla I.

TABLA II. PRUEBAS DE CONCEPTO

Protocolo	Prueba de Concepto
STP	<i>Claim root role</i>
CDP	Aprovechamiento de CDP
IEEE 802.1q	Salto de VLAN Aprovechamiento DTP
Telnet y SSH	Ataque de fuerza bruta
DHCP	Suplantación de servidor DHCP
ARP	ARP Posoning, <i>Man-in-the-middle</i>

Las herramientas para las pruebas son provistas por los instructores o docentes, entre las que destaca el framework Yersinia, que es un programa de software para realizar ataques (en nuestro caso pruebas de concepto) contra servicios y

protocolos de red tales como STP, CDP, DTP, DHCP, HSRP, 802.1Q y VTP. Estos ataques son realizados a nivel de capa 2.

El desarrollo del laboratorio se ejecuta en un contexto de tipo CtF (Capture the Flag) por equipos, con el propósito de promover la competencia y fortalecimiento de habilidades de concentración y trabajo en equipo.

En los escenarios planteados se asume que el atacante se encuentra en el mismo segmento de red que los dispositivos vulnerables, por lo que la evaluación y medición de los niveles de severidad son considerando este aspecto de ubicación del atacante.

Asimismo, es necesario considerar que los estudiantes al recibir entrenamiento sobre la metodología y el uso de herramientas de hacking están comprometidos con aspectos éticos y legales por lo que deberán firmar un Acuerdo de Compromiso sobre el uso adecuado de los conocimientos y herramientas a recibir. La finalidad es comprometer al estudiante a no usar el conocimiento adquirido para acciones fuera de la ley o ética. También se deberá emitir un “Disclaimer” sobre las técnicas mostradas.

En general las fases de desarrollo del Laboratorio son las siguientes:

- **Identificación de vulnerabilidad.** Este procedimiento es abierto a otras técnicas o formas de encontrar una vulnerabilidad, inclusive las herramientas (legales y éticas) a utilizar. El estudiante no queda limitado a una sola forma de proceder. Lo que se muestra en este trabajo es finalmente un ejemplo de un procedimiento clásico.
- **Verificación de vulnerabilidad.** Luego de identificar una vulnerabilidad se debe procede a comprobarla, es decir determinar si efectivamente es una vulnerabilidad que es susceptible de ser explotada o aprovechada. Las técnicas y herramientas son sugeridas, pero no quedan limitadas a esas.
- **Valoración CVSS.** En esta fase se debe estimar el grado de severidad de la vulnerabilidad que permite el ataque exitoso. Para la valoración se determina con el uso de la calculadora de CVSS versión 3.
- **Mitigación de la vulnerabilidad.** El estudiante deberá plantear las recomendaciones de mitigación asociadas a las soluciones que hayan determinado los fabricantes, sin embargo, no queda limitado a ese aspecto, por lo que el estudiante puede generar propuestas de mitigación diversas pero que deberá mostrar su eficiencia.
- **Alineación con los controles CIS.** En esta última fase el estudiante deberá analizar la correspondencia entre la vulnerabilidad de la prueba de concepto desarrollada con algún o algunos controles que CIS ha formulado.

A. Prueba de concepto de Claim Root Role

Sobre la topología base, se realizó la configuración de STP, definiendo a uno de los switches en el rol de “puente raíz” (root bridge) modificando la prioridad asociada. La Fig. 2 muestra el escenario.

La prueba de Claim Root Role consiste en la posibilidad de obtener el rol de Root para una topología conmutada con el

protocolo Spanning Tree activo. El efecto de este ataque es desestabilizar o modificar la topología STP de la red, lo cual puede llevar a errores de control de bucles de forma intermitente; así como aprovechar el rol de puente raíz (root bridge) para otros ataques como denegación de servicio y man-in-the-middle.

Para esta prueba de ataque se utilizó la herramienta Yersinia, y se intentó obtener el rol de root en la VLAN 20, asumiendo que un atacante malicioso logre conectarse físicamente a un puerto de switch en la VLAN 20. La valoración de la vulnerabilidad se muestra en la Tabla II.

Para mitigar el ataque se deben habilitar la siguiente configuración en los switches de la red:

```
S(config)#spanning-tree portfast bpduguard
```

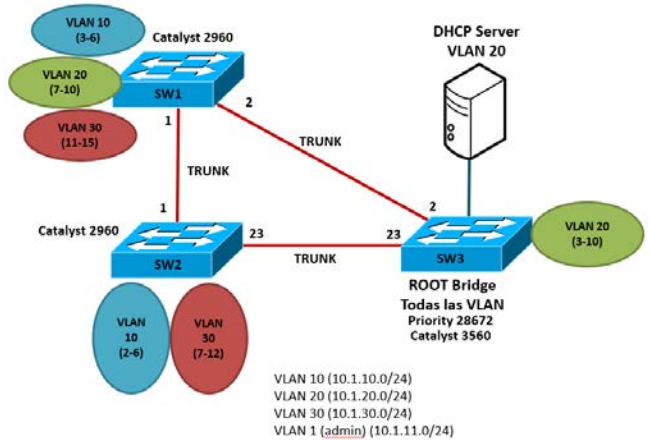


Fig. 2. Topología Pruebas de Concepto de Claim Root Role

TABLA III. VALORACIÓN CVSS DE CLAIM ROOT ROLE

Métrica Base	Valor
Vector de ataque	Adyacente
Complejidad de ataque	Bajo
Privilegios requeridos	Ninguno
Interacción con usuario	Ninguno
Alcance	Sin cambios
Confidencialidad	Ninguno
Integridad	Ninguno
Disponibilidad	Bajo
Puntuación Base CVSS v3.0	4.3

La Tabla III muestra el resultado de la valoración del nivel de severidad CVSS para el ataque de Claim Root Role que se establece en 4.3 como nivel medio, debido a que el vector de ataque es de modo adyacente, significa que se necesita estar en el mismo dominio de broadcast para explotar la vulnerabilidad, además de solo afectar a la disponibilidad de una forma baja.

B. Prueba de concepto de aprovechamiento de CDP

El protocolo CDP, propietario de Cisco, tiene la función de mantener informado a los equipos vecinos acerca del tipo de dispositivo conectado y sus características (marca, modelo, versión de IOS, etc.), por lo que esta información puede ser obtenida aprovechándose de la actividad del protocolo CDP

en las interfaces que no son necesarias como aquellas conectadas a dispositivos finales (PCs, laptops, impresoras, etc.). Se considera un ataque de reconocimiento.

Para realizar la prueba de concepto se utilizó la herramienta Yersinia en modo interactivo. En la Fig. 3 se puede observar la información obtenida a partir del envío de mensajes CDP. La información obtenida en el ejercicio es de un equipo Catalyst 2960, de 24 puertos FastEthernet, que incluye datos sobre el IOS o sistema operativo del equipo. La valoración de la vulnerabilidad se muestra en la Tabla III.

Para mitigar el aprovechamiento del protocolo de descubrimiento se debe deshabilitar los anuncios de CDP en las interfaces que no están conectados a equipos de red.

Para deshabilitar en una interfaz específica:

```
S(config-if)#no cdp enable
```

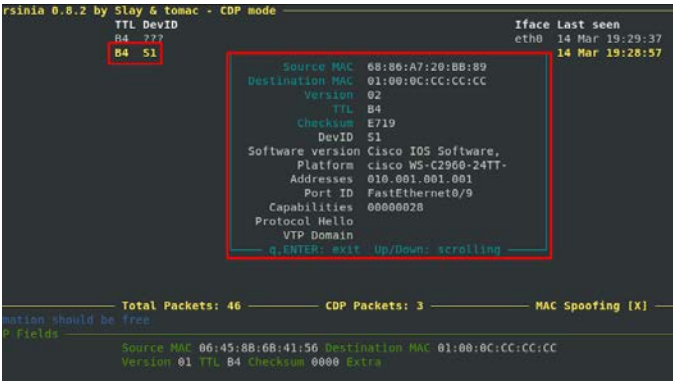


Fig. 3. Información obtenida de CDP

TABLA IV. VALORACIÓN CVSS DE APROVECHAMIENTO CDP

Métrica Base	Valor
Vector de ataque	Adyacente
Complejidad de ataque	Bajo
Privilegios requeridos	Ninguno
Interacción con usuario	Ninguno
Alcance	Sin cambios
Confidencialidad	Bajo
Integridad	Ninguno
Disponibilidad	Ninguno
Puntuación Base CVSS v3.0	4.3

La Tabla IV muestra el resultado de la valoración del nivel de severidad CVSS para la vulnerabilidad que permite el aprovechamiento CDP, el cual está establecido en 4.3, es decir un nivel medio, debido a que el vector de ataque es de modo adyacente, significa que se necesita estar en el mismo dominio de broadcast para que el ataque sea exitoso, además de solo afectar la confidencialidad de manera baja.

C. Prueba de concepto de ataque a DTP (Dynamic Trunk Protocol)

Para realizar la prueba de concepto para la explotación referida a VLAN en switches se ha establecido el escenario mostrado en la Fig. 4.

El Protocolo Troncal Dinámico (DTP por sus siglas en inglés) propietario de Cisco, tiene como función principal

negociar un enlace para convertirse en un enlace troncal o no. Para tal efecto, las interfaces deben tener habilitado la negociación dinámica. La vulnerabilidad radica en que la negociación está habilitada por defecto en las interfaces que conectan dispositivos finales, por lo que esto puede ser aprovechado por atacantes para establecer un enlace troncal no deseado, consiguiendo de esta manera al tráfico de todas las VLANs permitidas en los enlaces troncales.

Para realizar la prueba de concepto se utilizó la herramienta Yersinia.

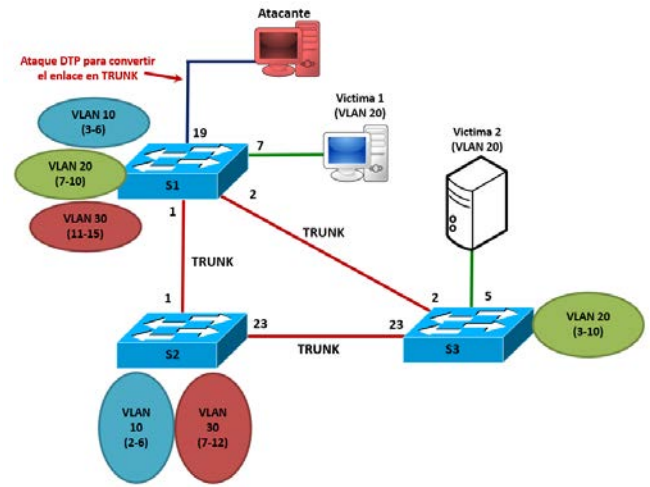


Fig. 4. Topología para Pruebas de Concepto de DTP

En la Fig. 5, se puede observar la negociación del protocolo DTP para lograr un enlace troncal con el equipo atacante.

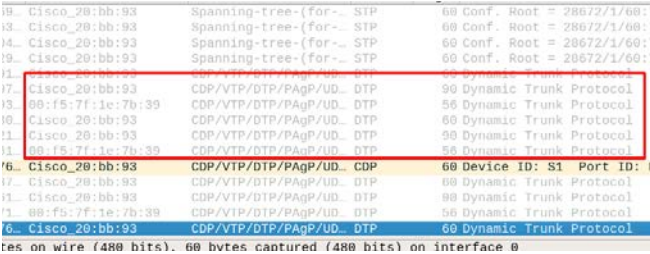


Fig. 5. Negociación DTP con Yersinia

Después de la negociación de enlace troncal, se puede verificar en el switch (S1) estableció un enlace troncal de forma automática. La Fig. 6 muestra la tabla de enlaces troncales en el switch S1, y la interfaz Fa0/19 se ha convertido en un enlace troncal, y esta interfaz es la que conecta a una PCs que es el equipo atacante.

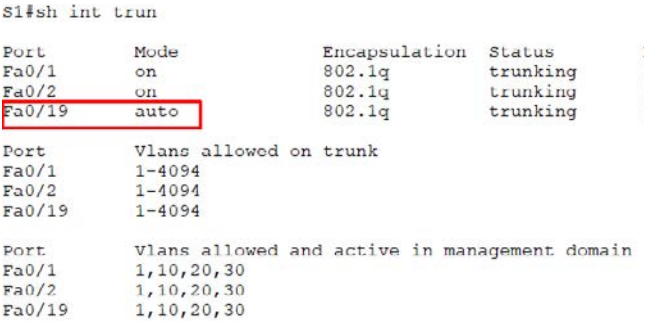


Fig. 6. Enlaces Troncales en el switch S1.

Asimismo, se puede visualizar en la Fig. 7, el tráfico capturado que pertenece a otras VLANs, en este caso de la VLAN 10 (10.1.10.0/24).

8...	fe80::9c57:1bbc:47d...ff02::1:3	ICMPv6	90	Multicast Listener Report
1...	fe80::9c57:1bbc:47d...ff02::1:3	LLMNR	88	Standard query 0x97ea AN
3...	10.1.10.20	224.0.0.252	LLMNR	88 Standard query 0x97ea AN
3...	fe80::9c57:1bbc:47d...ff02::1:3	ICMPv6	90	Multicast Listener Report
3...	fe80::9c57:1bbc:47d...ff02::1:3	LLMNR	86	Standard query 0xa790 A
3...	10.1.10.20	224.0.0.252	LLMNR	86 Standard query 0xa790 A
8...	fe80::9c57:1bbc:47d...ff02::1:3	ICMPv6	90	Multicast Listener Report
8...	fe80::9c57:1bbc:47d...ff02::1:3	ICMPv6	90	Multicast Listener Report
5...	10.1.10.20	10.1.10.255	NBNS	110 Registration NB USER-PC

Fig. 7. Tráfico de otras VLAN capturado

La Tabla V muestra el resultado de la valoración del nivel de severidad CVSS para la vulnerabilidad que permite el ataque a DTP, el cual está establecido en 4.3, es decir un nivel medio, debido a que el vector de ataque es de modo adyacente, significa que se necesita estar en el mismo dominio de broadcast para que el ataque sea exitoso, además de solo afectar la confidencialidad de manera baja.

TABLA V. VALORACIÓN CVSS DE ATAQUE A DTP

Métrica Base	Valor
Vector de ataque	Adyacente
Complejidad de ataque	Bajo
Privilegios requeridos	Ninguno
Interacción con usuario	Ninguno
Alcance	Sin cambios
Confidencialidad	Bajo
Integridad	Ninguno
Disponibilidad	Ninguno
Puntuación Base CVSS v3.0	4.3

Para evitar la negociación de un enlace troncal en interfaces no deseadas, se debe suprimir la negociación en la interfaz específica, de la siguiente forma:

```
S(config-if)# switchport mode access | trunk
S(config-if)# switchport nonegotiate
```

D. DHCP Spoofing

La prueba de concepto para DHCP Spoofing requiere de un servidor legítimo en la red experimental construida para las pruebas anteriores. La Fig. 8 muestra el escenario utilizado para el tipo de ataque de suplantación de servidor DHCP.

La Fig. 9 y Fig. 10 muestran los resultados de la prueba, utilizando como herramienta el módulo de Metasploit Framework auxiliary/server/dhcp.

Los hosts de la red local que tienen habilitado la configuración de sus direcciones IP mediante DHCP, recibiendo los parámetros falsos del servidor DHCP Rogue.

La Tabla VI muestra el resultado de la valoración del nivel de severidad CVSS para la vulnerabilidad que permite el aprovechamiento CDP, el cual está establecido en 5.3, es decir un nivel medio, debido a que el vector de ataque es de modo adyacente, significa que se necesita estar en el mismo dominio de broadcast para que el ataque sea exitoso, además de afectar la confidencialidad y disponibilidad de manera baja.

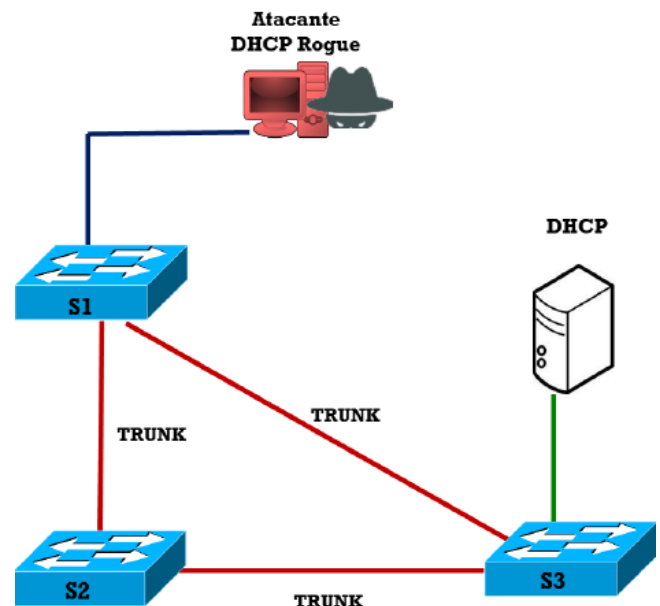


Fig. 8. Topología para Prueba de Concepto de DHCP Spoofing

```
msf auxiliary(server/dhcp) > show options

Module options (auxiliary/server/dhcp):

Name          Current Setting  Required  Description
-----
BROADCAST     192.168.20.150  no        The broadcast address to
DHCIPIPEND    192.168.20.100  no        The last IP to give out
DHCIPISTART   192.168.20.100  no        The first IP to give out
DNSSERVER     192.168.100.100 no        The DNS server IP address
DOMAINNAME    malosocom       no        The optional domain name
FILENAME      malosocom       no        The optional filename of
HOSTNAME      malosocom       no        The optional hostname to
HOSTSTART     malosocom       no        The optional host integer
NETMASK       255.255.255.0   yes       The netmask of the local
ROUTER        192.168.20.1    no        The router IP address
SRVHOST       10.1.20.200     yes       The IP of the DHCP server
```

Fig. 9. Módulo de Metasploit auxiliary/server/dhcp

```
jefatutalabredes@jefatutalabredes:~$ ifconfig
enp5s0 Link encap:Ethernet direcciónHW f0:bf:97:69:0a:1e
      Direc. inet:192.168.20.101 Másc:192.168.20.255 Másc
      Dirección inet6: fe80::9326:1f22:ebc0:6888/64 Alcance:El
      ACTIVO DIFUSIÓN FUNCIONANDO MULTICAST MTU:1500 Métrica:1
      Paquetes RX:87 errores:0 perdidos:0 overruns:0 frame:0
      Paquetes TX:132 errores:0 perdidos:0 overruns:0 carrier
      colisiones:0 long.colatX:1000
      Bytes RX:8134 (8.1 KB) TX bytes:17018 (17.0 KB)

lo Link encap:Bucle local
  Direc. inet:127.0.0.1 Másc:255.0.0.0
  Dirección inet6: ::1/128 Alcance:Anfitrión
  ACTIVO BUCLE FUNCIONANDO MTU:65536 Métrica:1
  Paquetes RX:419 errores:0 perdidos:0 overruns:0 frame:0
  Paquetes TX:419 errores:0 perdidos:0 overruns:0 carrier
  colisiones:0 long.colatX:1
  Bytes RX:31831 (31.8 KB) TX bytes:31831 (31.8 KB)

jefatutalabredes@jefatutalabredes:~$ cat /etc/resolv.conf
# Dynamic resolv.conf(5) file for glibc resolver(3) generated by
# DO NOT EDIT THIS FILE BY HAND -- YOUR CHANGES WILL BE OVERW
nameserver 192.168.100.100
nameserver 127.0.1.1
search malosocom
```

Fig. 10. Módulo de Metasploit auxiliary/server/dhcp

Para la mitigación se debe habilitar port-security y DHCP snooping en los switches para evitar el procesamiento y reenvío de mensajes DHCP (DHCP OFFER, DHCP DISCOVER, DHCP REQUEST y DHCP ACK) por los puertos de switch que no son de confianza; es decir por aquellos puertos donde no debería estar conectado un servidor DHCP.

TABLA VI. VALORACIÓN CVSS DE PRUEBA DE DCHP SPOOFING

Métrica Base	Valor
Vector de ataque	Adyacente
Complejidad de ataque	Bajo
Privilegios requeridos	Ninguno
Interacción con usuario	Ninguno
Alcance	Sin cambios
Confidencialidad	Bajo
Integridad	Ninguno
Disponibilidad	Bajo
Puntuación Base CVSS v3.0	5.4

E. ARP Poisoning – Main-in-the-middle

Un ataque de hombre en el medio a nivel de capa 2, requiere de un ataque previo como el de envenenamiento de ARP (ARP Poisoning) para que sea efectivo. La Fig. 11 muestra la topología utilizada para este tipo de prueba.

Las herramientas utilizadas para esta prueba fueron Ettercap para el envenenamiento de las tablas ARP de los hosts víctimas (ver Fig. 12) y el Wireshark para la verificación del ataque. La valoración de la vulnerabilidad se muestra en la Tabla VI.

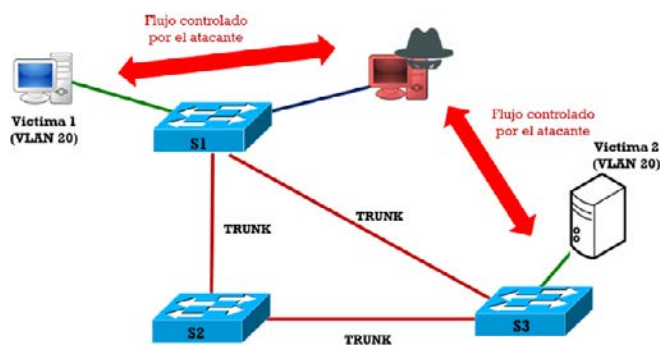


Fig. 11. Topología para la prueba de concepto de ARP Poisoning

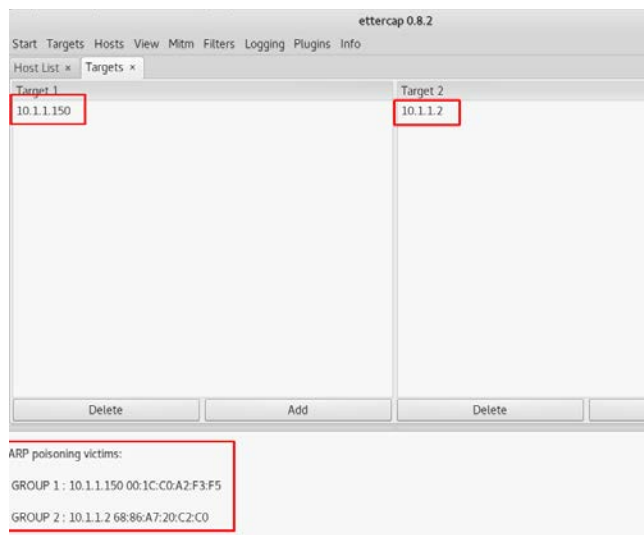


Fig. 12. Envenenamiento de ARP con Ettercap

TABLA VII. VALORACIÓN CVSS DE PRUEBA DE ARP SPOOFING

Métrica Base	Valor
Vector de ataque	Adyacente
Complejidad de ataque	Bajo
Privilegios requeridos	Ninguno
Interacción con usuario	Ninguno
Alcance	Sin cambios
Confidencialidad	Alto
Integridad	Ninguno
Disponibilidad	Ninguno
Puntuación Base CVSS v3.0	6.5

La Tabla VII muestra el resultado de la valoración del nivel de severidad CVSS para la vulnerabilidad que permite el al ataque de ARP Spoofing el cual está establecido en 6.5, es decir un nivel medio, debido a que el vector de ataque es de modo adyacente, significa que se necesita estar en el mismo dominio de broadcast para que el ataque sea exitoso, además de afectar la confidencialidad de forma alta, porque es posible la captura de mensajes eventualmente no cifrados que pueden contener información sensible o confidencial.

Para mitigar los ataques por envenenamiento de ARP, se debe habilitar el mecanismo de inspección dinámica de ARP en los puertos del switch

F. Ataque de diccionario Telnet y SSH

Para las pruebas de autenticación con protocolos de acceso remoto como TELNET y SSH, se realizaron ataques de diccionario. Como se muestra en la Fig. 13, los equipos admiten gran cantidad de intentos para autenticarse, lo que resulta una debilidad que puede ser aprovechada por agentes maliciosos al utilizar diccionarios que pueden ser elaborados de forma personalizada y ser exitosos al encontrar una credencial válida.

```
root@kali:~# hydra -s 23 -v -P Switches-Attack/PassCisco 10.1.1.1
Hydra v8.6 (c) 2017 by van Hauser/THC - Please do not use in military or se
for illegal purposes.

Hydra (http://www.thc.org/thc-hydra) starting at 2018-03-14 16:31:26
[WARNING] you should set the number of parallel task to 4 for cisco service
[WARNING] Restorefile (you have 10 seconds to abort... (use option -I to sh
ssion found, to prevent overwriting, ./hydra.restore
[DATA] max 16 tasks per 1 server, overall 16 tasks, 16 login tries (l:1/p:1
[DATA] attacking cisco://10.1.1.1:23/
[VERBOSE] Resolving addresses ... [VERBOSE] resolving done
[STATUS] attack finished for 10.1.1.1 (waiting for children to complete tes
[23][cisco] host: 10.1.1.1 password: Cisco1234
1 of 1 target successfully completed, 1 valid password found
[WARNING] Writing restore file because 1 final worker threads did not compl
[ERROR] 1 target did not resolve or could not be connected
[ERROR] 16 targets did not complete
```

Fig. 13. Ataque de diccionario Telnet

La herramienta utilizada para esta prueba fue Hydra con un diccionario elaborado con palabras comunes que corresponden a contraseñas para este tipo de dispositivos.

La Tabla VIII muestra el resultado de la valoración del nivel de severidad CVSS para la vulnerabilidad que permite el ataque de diccionario contra servicios de Telnet y SSH, el cual está establecido en 5.3, es decir un nivel medio, debido a que el vector de ataque es a nivel de red, significa que es posible realizar el ataque de forma remota desde otros segmentos de red, además de afectar solo la confidencialidad de manera baja.

TABLA VIII. VALORACIÓN CVSS DE PRUEBA DE ATAQUE DE DICCIONARIO CONTRA SERVICIOS TELNET Y SSH

Métrica Base	Valor
Vector de ataque	Red
Complejidad de ataque	Bajo
Privilegios requeridos	Ninguno
Interacción con usuario	Ninguno
Alcance	Sin cambios
Confidencialidad	Bajo
Integridad	Ninguno
Disponibilidad	Ninguno
Puntuación Base CVSS v3.0	5.3

Para mitigar esta vulnerabilidad se recomienda limitar el origen de las conexiones con listas de control de acceso (ACL).

G. Alineación con los Controles CIS

De acuerdo con la rúbrica formulada, el estudiante deberá asociar las pruebas de concepto con los controles que correspondan a los 20 controles críticos publicado por el Center for Internet Security (CIS). Un ejemplo de esta correspondencia se puede apreciar en la Tabla IX, donde se observa que se alinean principalmente con dos sub-controles correspondientes al control 11 de CIS “Configuración segura de los equipos de red, tales como cortafuegos, enrutadores y conmutadores”.

TABLA IX. ALINEACIÓN CON LOS CONTROLES DE CIS

Prueba de concepto	Sub Control	Descripción
Claim root role	11.1. Mantener configuraciones de seguridad estandarizadas en equipos de red	Mantenga configuración de seguridad estandarizadas y documentados para todos los equipos de red autorizados
Suplantación de servidor DHCP		
ARP Poisoning, Man-in-the-middle		
Aprovechamiento de CDP		
Salto de VLAN Aprovechamiento DTP	11.2. Documentar las reglas de configuración de tráfico	Todas las reglas de configuración que permiten que el tráfico fluya a través de dispositivos de red deben documentarse en un sistema de gestión de configuración con un fin de negocio específico para cada regla, el nombre de un individuo específico responsable de esa necesidad de negocio y una duración esperada de la necesidad
Ataque de fuerza bruta		

H. Rúbrica para el desarrollo del Laboratorio

La rúbrica formulada es la que se muestra en la Tabla X que abarca las fases del Laboratorio propuesto.

TABLA X. RÚBRICA PARA DESARROLLO DEL LABORATORIO

Criterio	Nivel
Identificación de vulnerabilidad (reconocimiento)	(4) <i>Excelente.</i> Identifica con éxito la vulnerabilidad con el apoyo de herramientas adecuadas o de forma manual con previa detección de la presencia de operación de protocolo y/o puerto asociado.
	(3) <i>Satisfactorio</i> Puede identificar la vulnerabilidad con el apoyo de herramientas adecuadas o de forma manual con previa detección de la presencia de operación de protocolo y/o puerto asociado.
	(2) <i>Puede Mejorar</i> Ocasionalmente puede identificar la vulnerabilidad con el apoyo de herramientas adecuadas o de forma manual con previa detección de la presencia de operación de protocolo y/o puerto asociado.
	(1) <i>Inadecuado (necesita ayuda)</i> Necesita asistencia para identificar la vulnerabilidad con uso de herramientas adecuadas o de forma manual con previa detección de la presencia de operación de protocolo y/o puerto asociado
Verificación de vulnerabilidad	(4) <i>Excelente.</i> Verifica con éxito la presencia de la vulnerabilidad con el apoyo de herramientas seleccionadas adecuadamente.
	(3) <i>Satisfactorio</i> Puede verificar la presencia de la vulnerabilidad con el apoyo de herramientas seleccionadas adecuadamente.
	(2) <i>Puede Mejorar</i> Ocasionalmente puede verificar la presencia de la vulnerabilidad con el apoyo de herramientas seleccionadas adecuadamente
	(1) <i>Inadecuado (necesita ayuda)</i> Necesita asistencia para verificar la presencia de la vulnerabilidad con el apoyo de herramientas seleccionadas adecuadamente
Valoración CVSS	(4) <i>Excelente.</i> Realiza con éxito la valoración de cada vulnerabilidad verificada con las métricas base de CVSS v3 y considerando las características de las pruebas de concepto.
	(3) <i>Satisfactorio</i> Puede realizar la valoración de cada vulnerabilidad verificada con las métricas base de CVSS v3 y considerando las características de las pruebas de concepto.
	(2) <i>Puede Mejorar</i> Ocasionalmente puede realizar la valoración de cada vulnerabilidad verificada con las métricas base de CVSS v3 y considerando las características de las pruebas de concepto.
	(1) <i>Inadecuado (necesita ayuda)</i> Necesita asistencia para realizar la valoración de cada vulnerabilidad verificada con las métricas base de CVSS v3 y considerando las características de las pruebas de concepto.
Mitigación de la vulnerabilidad	(4) <i>Excelente.</i> Identifica con éxito la solución para mitigación y describe correctamente el procedimiento de la aplicación de las recomendaciones técnicas asociadas a la solución de mitigación.
	(3) <i>Satisfactorio</i> Puede identificar la solución para mitigación y describe correctamente el procedimiento de la aplicación de las recomendaciones técnicas asociadas a la solución de mitigación.
	(2) <i>Puede Mejorar</i> Ocasionalmente identifica la solución para mitigación y describe el procedimiento de la aplicación de las recomendaciones técnicas asociadas a la solución de mitigación.
	(1) <i>Inadecuado (necesita ayuda)</i>

	Necesita asistencia para identificar la solución de mitigación y describir correctamente el procedimiento de la aplicación de las recomendaciones técnicas asociadas a la solución de mitigación.
Alineación con controles CIS	(4) <i>Excelente</i> . Establece con éxito la correspondencia entre la vulnerabilidad y algún control o controles de CIS.
	(3) <i>Satisfactorio</i> Puede establecer la correspondencia entre la vulnerabilidad y algún control o controles de CIS
	(2) <i>Puede Mejorar</i> Ocasionalmente establece la correspondencia entre la vulnerabilidad y algún control o controles de CIS
	(1) <i>Inadecuado (necesita ayuda)</i> Necesita asistencia para establecer la correspondencia entre la vulnerabilidad y algún control o controles de CIS

IV. DISCUSIÓN

El Laboratorio propuesto es una forma efectiva de fortalecer y apoyar el proceso de enseñanza-aprendizaje de la seguridad de redes, específicamente en infraestructuras basadas en switches administrables.

Las pruebas de concepto seleccionadas y las vulnerabilidades comunes son básicas, que en entornos de red de producción maduros probablemente no están presentes, sin embargo, son vulnerabilidades aún vigentes y vectores de ataques válidos, además de ser una fuente de aprendizaje importante para los estudiantes de carreras afines.

En trabajos futuros es posible ampliar las características de las pruebas de concepto utilizando equipos de diversos fabricantes y controles adicionales como los contemplados en la ISO/IEC 27001, PCI, entre otros, de tal forma que el estudiante tenga un panorama más completo de los aspectos técnicos y de gestión de la seguridad informática; asimismo es deseable pasar a niveles más sofisticados de pruebas de concepto como la manipulación de mecanismos de criptografía, o ataques de Denegación de Servicio (DoS).

Las herramientas utilizadas no son las únicas con las que se puede desarrollar los laboratorios, sin embargo, se recomienda analizar cómo opera cada herramienta que se pretenda usar. El estudiante puede elegir distintas herramientas a las establecidas para el laboratorio, lo que puede permitir desarrollar capacidades de análisis complementarias.

Las recomendaciones de mitigación o corrección de las vulnerabilidades pueden ser tomadas de las indicaciones de los fabricantes, pero también pueden ser propuestas por el estudiante mostrando su eficiencia. De esta manera se deja abierto a las posibilidades de generar soluciones diversas.

V. CONCLUSIONES

El Laboratorio propuesto como una forma de soporte y apoyo al proceso de enseñanza-aprendizaje de la seguridad de redes a nivel universitario ofrece un paquete práctico que incluye las fases generales y comunes de un ataque (seguridad ofensiva), los procedimientos de mitigación (seguridad defensiva) y los controles de seguridad basados en una referencia válida como CIS (gestión de la seguridad informática). Este laboratorio, aunque limitado a un solo fabricante y tipo de control, es útil y motivador.

AGRADECIMIENTOS

Al Laboratorio de Redes y Seguridad y a la Facultad de Ingeniería en Informática y Sistemas de la Universidad Nacional Agraria de la Selva.

REFERENCIAS

- [1] M. A. Rahman y E. Al-Shaer, «A declarative approach for global network security configuration verification and evaluation,» de *12th IFIP/IEEE International Symposium on Integrated Network Management*, 2011.
- [2] ISO/IEC, «ISO/IEC 27002 -Tecnología de la Información – Técnicas de seguridad – Código para la práctica de la gestión de la seguridad de la información,» ISO, 2013.
- [3] J. Cioara, D. Minutella y H. & Stevenson, Exam Prep CCNA 640-802., Indianapolis: Pearson Education, 2008.
- [4] W. K. Alzubaidi, L. Cai y S. A. Alyawer, «A New Verification Method To Prevent Security Threads Of Unsolicited Message In Ip Over Ethernet Networks,» *International Journal of Computer Networks & Communications*, vol. 4, n° 6, pp. 21-31, 2012.
- [5] The MITRE Corporation., «CVE - Common Vulnerabilities and Exposures,» 2019. [En línea]. Available: <https://cve.mitre.org/>. [Último acceso: noviembre 2018].
- [6] O. Santos y J. & Stuppi, CCNA Security 210-260, Indianapolis: Pearson Education - CiscoPress, 2015.
- [7] FIRST.org, Inc, «Common Vulnerability Scoring System v3.1: Specification Document,» 2019. [En línea]. Available: <https://www.first.org/cvss/v3.1/specification-document>. [Último acceso: enero 2019].
- [8] Center for Internet Security, «Center for Internet Security,» 2019. [En línea]. Available: <https://www.cisecurity.org>. [Último acceso: Diciembre 2018].
- [9] W. R. Marchand, E. Vega y J. Santillan, «Capture the Flag for Computer Security Learning,» de *IX Congreso Iberoamericano de Seguridad Informática y IV Taller Educativo TIBETS*, Buenos Aires, 2017.
- [10] A. Bechtsoudis y N. Sklavos, «Aiming at Higher Network Security Through Extensive Penetration Tests,» *IEEE Latin America Transactions*, vol. 10, n° 3, pp. 1752-1756, 2012.
- [11] D. Shmaryahu, «Constructing Plan Trees for Simulated Penetration Testing,» de *The 26th International Conference on Automated Planning and Scheduling*, London, 2016.
- [12] A. Tewai y A. Kumar Misra, «Evaluation and Taxonomy of Penetration Testing,» *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 3, n° 8, pp. 5297 - 5302, 2015.
- [13] Cisco System, CCNA Routing & Switching, California: Cisco Press, 2016.
- [14] Y. Bhajji, «Understanding, Preventing, and Defending Against Layer 2 Attacks,» 2009. [En línea]. Available: https://www.cisco.com/c/dam/global/en_ae/assets/exposaudi2009/assets/docs/layer2-attacks-and-mitigation-t.pdf. [Último acceso: Octubre 2018].
- [15] G. Weidman, Penetration Testing. A Hands-on Introduction to Hacking, San Francisco, California: Law, 2014.
- [16] N. Jaswal, Mastering Metasploit, Birmingham, Mumbai: Packt Publishing, 2015.
- [17] C. McNab, Network Security Assessment, San Francisco: O'Reilly Media, 2015.
- [18] P. Gonzáles Pérez, Metasploit para pentester, Madrid: 0xWord, 2014.
- [19] T. Kiravuo, M. Sarela y J. Manner, «A Survey of Ethernet LAN Security,» *IEEE COMMUNICATIONS SURVEYS & TUTORIALS*, vol. 15, n° 3, p. 1477–1491, 2013.

William-Rogelio Marchand-Niño, Ingeniero de Sistemas otorgado por la Universidad Nacional del Centro del Perú, con maestría en Dirección Estratégica de TI de la Universidad de Piura, con 18 años de experiencia académica en UNAS, UDH, UPLA. Desde el año 2004 es profesor asociado en la UNAS. Ha impartido más de 90 cursos de pregrado en diferentes universidades. Instructor CISCO por 12 años. Posee múltiples certificaciones de la Industria como PMP, ITIL Foundation, CCNA, MTA. Director del Centro de Tecnologías de Información y Comunicación de la Universidad Nacional Agraria de la Selva. Miembro Senior de la IEEE.

José Martín Santillán Ruiz, Ingeniero en Informática y Sistemas de la Universidad Nacional Agraria de la Selva. Profesor auxiliar en la UNAS, con mas de 7 años de experiencia en docencia. Miembro IEEE. Coordinador del área de Gestión de la Red Corporativa en el Centro de Tecnologías de la Información y Comunicación de la UNAS-Perú en el año 2016-2017.

Estrategias para la implementación y adopción de componentes curriculares de ciberseguridad en programas de grado y posgrado

F.A. Corredor¹, D. C. Franco², J. E. Martínez³.

Resumen - This paper presents the approach and application of strategies to assume two big problematic aspects of the cybersecurity sector, in which higher education institutions together with other stakeholders (productive sector and government) should cooperate and establish appropriate solutions. As a first aspect, the curricular design for integral training in cybersecurity is assumed, due to the density of themes, references of criteria and competences, which contrasts with the restrictions of space in the curricula of degree programs in information technologies (I.T.) with which you must address issues of foundation, deepening and emerging. moreover, the industry demand of the cybersecurity professionals, which currently exceeds supply; generating a global risk in organizations. The strategy is being implemented in a Colombian University, involving actively the student body to capture their interest in the area, as an objective of integral training and a graduated that generates labor supply.

Palabras Clave - Cybersecurity education, systems engineering, curricular strategies, undergraduate computing programs.

informática o computación, como son la computación inteligente, big data y analítica, cloud computing, IoT y ciberseguridad, que conllevan a tomar decisiones de cobertura a nivel curricular para dejar componentes en el núcleo de formación – obligatorio – de las planes de estudio y dejar otros desde la oferta de electivas.

En el caso de ciberseguridad; se debe ser cuidadoso en la forma de estructurar y organizar la gran densidad de sus ejes temáticos, para responder al contexto local y global, pero también al proceso pedagógico que capte la atención de estudiantes, motivándolos hacia el área de la ciberseguridad (seguir formándose y ejercer laboralmente) y contribuir a la disminución del déficit de profesionales en ciberseguridad que sigue en aumento.

El artículo presenta los aspectos que definen el problema y propone una estrategia que está siendo adoptada en una universidad pública en Colombia, posteriormente una discusión de los aspectos encontrados y finalmente las conclusiones.

I. INTRODUCCIÓN

LA ciberseguridad es un aspecto que ha sido reconocido como elemento estratégico a nivel organizacional, abordado desde los diferentes estándares (CSF NIST, COBIT, ISO27001, ISO 27103, entre otros) y herramientas de software para gestión y operación específica, pero aunque hay cierto grado de consenso en los aspectos disciplinares propios de los contenidos tecnológicos, se genera cierto desacuerdo en los aspectos metodológicos y las temáticas a desarrollar, por las limitantes de espacio en los currículos de los programas profesionales de tecnologías de información (T.I.).

Desde la academia, se debe considerar esa gran cantidad de ejes temáticos y enfoques, así como la oferta de programas no formales y certificaciones, lo que dificulta garantizar la calidad y la secuencia organizada en ciclos y niveles de formación. Los diseños microcurriculares deben velar por desarrollar una gama de competencias (definidas por ABET, ABET CSAB, ACOFI, CSEC) las cuales deben cubrir aspectos generales y específicos de áreas fundamentales como la algoritmia y lógica, matemáticas, bases de datos, comunicaciones, etc. Pero también temas más avanzados que son de relevancia para los currículos de ingeniería de sistemas,

II. CONTEXTO DEL PROBLEMA

La amplia dependencia tecnológica de las sociedades y sectores económicos ha venido presentando una preocupante situación de riesgos a nivel de ciberseguridad, donde el eslabón más débil sigue siendo el recurso humano interno de las organizaciones (menos del 40% de las organizaciones realiza capacitación y concienciación del personal [1]) y la principal amenaza la generan en su mayoría atacantes humanos externos (69% de los ataques son generados por outsiders y el 52% de los ataques involucra operaciones de hacking [2]) lo cual dificulta las operaciones de protección y lo convierte en un problema que continúa en aumento. Esto ha generado la demanda de personal calificado para implementar una serie de factores de atención, gestión y operación, que garanticen su adecuado funcionamiento, dar cumplimiento a legislación específica y estar preparados ante las amenazas y vulnerabilidades que puedan materializar el riesgo. Esto acarrea varias situaciones de consideración para las instituciones de educación superior:

- El poco espacio de los currículos para asumir todos los temas emergentes y las competencias básicas, en el núcleo obligatorio de los planes de estudio. Lo que

¹ Felipe Andrés Corredor, M.Sc., Universidad de los Llanos, Grupo de Investigación GITECX, felcorredor@unillanos.edu.co.

² Diana Franco Mora, M.Sc., Universidad de los Llanos, Grupo de Investigación GITECX, dfranco@unillanos.edu.co.

³ Javier Eduardo Martínez, Mg., Universidad de los Llanos, Grupo de Investigación GITECX, jmartinez@unillanos.edu.co.

implica tomar decisiones y acciones respecto a las estrategias para formar integralmente a los nuevos profesionales con los conocimientos y capacidades suficientes en ciberseguridad.

- Incrementar la oferta de profesionales en ciberseguridad, ante la creciente demanda, que en la actualidad presenta un déficit de tres millones de profesionales de ciberseguridad en el mundo; lo que implica que deben plantearse estrategias (incluso extracurriculares) que incentiven el interés y agrado por esta disciplina.

En ambos casos, el sector educativo debe disponer de personal con amplio nivel de conocimiento teórico y aplicado, con visión analítica, crítica y conocimiento del contexto local y global. Lo cual conlleva a unir esfuerzos y capacidades de los profesionales del sector educativo con los profesionales del sector productivo [3]. Y aunque los tiempos de la academia son difíciles de sincronizar con los tiempos que maneja la industria, deben establecerse canales de cooperación y trabajo colaborativo que permitan mantener esa retroalimentación para los respectivos propósitos. (Formación, investigación y proyección social, así como la continuidad del negocio, innovación y oferta de servicios). Los proyectos de investigación, pasantías, cursos de formación bajo demanda, organización de eventos de concienciación, son algunas estrategias muy efectivas de propósitos comunes en ambos sectores.

En el primer aspecto, “diseño curricular” necesario, apropiado, y suficiente; las instituciones diseñan y adoptan una serie de estrategias para abordar una formación integral en ciberseguridad [4], en la cual se pueden definir y asumir los siguientes componentes:

- Teóricos: Matemáticas, probabilidad, estadística, grafos, teoría de la información, criptografía (sustitución, permutación, difusión, confusión, cifra simétrica en bloque y en flujo, cifra asimétrica, funciones hash, intercambio de clave, firma digital) y complejidad algorítmica.
- Tecnológicos: Programación, Sistemas operativos, sistemas distribuidos, Redes y comunicaciones, sistemas web y servicios telemáticos, arquitectura de software e infraestructura, etc. y sus respectivas habilidades tecnológicas.
- Organizacionales: Estándares y buenas prácticas de ISO, IEEE, NIST, OWASP, etc. Sistemas de gestión (procesos, riesgos, ciberseguridad), gobernanza de T.I., etc.
- Éticos y legales. Códigos de ética (Copnia – ley 842 de 2003, Código de ética de ACM), Ley 1273 de 2009, Ley 1928 de 2018 - Acuerdo de Budapest, Ley 1581 de 2012 (Protección de dato personales), Ley 1712 de 2014 (Transparencia de la información pública).
- Habilidades blandas: liderazgo, trabajo en equipo, comunicación verbal y escrita, argumentación, resolución de problemas, gestión del tiempo, etc. [5]

En una amplia gama de enfoques que provee la disciplina de ciberseguridad se hace necesario estructurar una secuencia adecuada de ejes temáticos en diferentes niveles de formación profesional y tomando como referentes los lineamientos curriculares internacionales, provistos por ACM-IEEE desde los enfoques de la computación, pero principalmente desde la

disciplina de la ciberseguridad establecida en el CSEC 2017 – Cybersecurity “*Curriculum Guidelines for Post-Secondary Degree Programs in Cybersecurity*”, que establece las características de un programa de ciberseguridad y sus componentes estructurales del currículo como son : seguridad de datos, seguridad de software, seguridad de componentes, seguridad en conexiones, seguridad en sistemas y seguridad en el recurso humano [6]. Lo cual es completamente compatible con la perspectiva asumida en cada uno de los elementos planteados anteriormente para formación integral en el programa de ingeniería de sistemas de la Universidad de los Llanos; cuyo diseño se basó en un enfoque híbrido entre el lineamiento curricular en tecnologías de información – TI.2017 [5] y Ciencias de la computación CS.2013 [7] de ACM-IEEE.

TABLA I.
ASPECTOS CURRICULARES DE REFERENCIA, ACM-IEEE

Enfoque	Áreas o aspectos de referencia
Tecnologías de Información. T.I. 2017	ITE-CSP Cybersecurity Principles (6% del currículo obligatorio) y ITS-CEC Cybersecurity Emerging Challenges (4% complementario)
Ciencias de la computación. C.S. 2013	IAS (Information Assurance and Security) Es una de las 18 áreas de conocimiento en el núcleo del plan de estudios; Dispone de seis horas en el núcleo y 31,5 horas de IAS distribuidas en otras áreas.
Ciberseguridad. CSEC 2017	Ciberseguridad como disciplina y áreas de la Seguridad: datos, software, componentes, conexiones, sistemas y humano, así como la perspectiva de la industria.

La Universidad de los Llanos, desde la Facultad de Ciencias Básicas e Ingeniería cuenta con programas formales de grado y posgrado, así como de programas no formales como diplomados, en los cuales, se ha logrado introducir el componente de ciberseguridad, manteniendo una línea secuencial de temáticas y niveles de profundidad.

TABLA II.
CRÉDITOS EN CIBERSEGURIDAD EN PROGRAMAS DE T.I. EN LA UNIVERSIDAD DE LOS LLANOS

Programa	Nivel	Créditos total	Créditos. Relación. Cibersegur	Aval MEN
Arquitectura de software y seguridad	Diplomado	96 Horas	48 Horas	No requerido (2019)
Ingeniería de Sistemas	Grado	165	8	Alta Calidad (2016) por 6 años.
Ingeniería Electrónica	Grado	169	4	Registro Calificado (2017) por 7 años
Ingeniería de Software	Especialización	28	3	Registro calificado (2014) por 7 años.
Tecnologías Digitales emergentes	Maestría, En construcción.	52	5	En Construcción, actualización al decreto 1330 del MEN

A nivel de pregrado, ciberseguridad se desarrollará a través de un curso obligatorio, nuevo, en séptimo semestre denominado “seguridad de la información” y un curso de profundización en décimo semestre denominado Curso de Profundización III – “aspectos avanzados de ciberseguridad” del área de teleinformática; plan de estudios que se encuentra acreditado

de alta calidad por el MEN Colombia para un periodo de seis años.



Figura 1. Estudiantes Décimo semestre, Curso de Seguridad de información, Ingeniería de sistemas.

A nivel de posgrado, desde la especialización en Ingeniería del software, se oferta una electiva denominada “ciberseguridad y hacking ético”, la cual tiene un enfoque ofensivo, donde se presentan aspectos de gestión y se sigue un proceso metodológico de penetration testing (PTES – Penetration Testing Execution Estándar) [8], hasta la entrega de reporte y plan de acción. Así mismo se presentó una electiva en un programa de maestría con aspectos avanzados de criptografía y aplicaciones pasando por PKI y blockchain, hasta modelos de seguridad y arquitectura de seguridad. Estos cursos de posgrado pueden ser tomados por estudiantes de Ingeniería de sistemas como opción de grado.

A nivel de educación no formal, se diseñó y desarrolló un diplomado denominado “Arquitectura de software y seguridad”, que está estructurado con el enfoque de seguridad para aplicaciones web, así como la arquitectura de infraestructura y la arquitectura de software, tomando como referente principal la guía de pruebas de OWASP 4.0 (OTG4) [9].

Para el segundo aspecto problemático, planteado en este artículo, que implica el déficit de profesionales en ciberseguridad, se requiere trabajar anticipadamente para que el estudiante mantenga el interés en la disciplina después del grado profesional, para ejercer en el área y continuar capacitándose (preferiblemente con incentivos financieros de apoyo). Diferentes organizaciones y el gobierno han determinado las situaciones de contexto que exponen varios de los aspectos problemáticos y las oportunidades que se tienen desde las instituciones que forman para la ciberseguridad:

El CONPES 3854 de 2016 que define “la política de seguridad digital - Colombiana” plantea la necesidad de formación para sus estudiantes en todos los niveles educativos y también para los profesores Universitarios, como se indica literalmente “Solicitar al Ministerio de Educación Nacional: a. Crear contenidos educativos complementarios relacionados con la gestión de riesgos de seguridad digital, y capacitar a: (i) los estudiantes de educación básica y media, (ii) los estudiantes de educación superior, y (iii) a los docentes” [10]; labor que han liderado desde el Ministerio de las TIC (MINTIC) y las Universidades Colombianas. Sin embargo, esto contrasta con los resultados emitidos por la encuesta nacional de seguridad informática de 2019 realizada

por la asociación Colombiana de Ingenieros de sistemas (ACIS) a través de la cual concluye que solo el 28% de los empresarios reconoce que se están ofreciendo programas de grado y posgrado formales en ciberseguridad y el 31% considera que no hay suficiente investigación en el área [11].

A esto se suma que a nivel mundial existe un gran déficit de profesionales en ciberseguridad, según lo plantea la Organización de Estados Americanos (OEA) y que afecta seriamente la adopción de su programa de ciberseguridad por parte de las organizaciones y los gobiernos del continente [12] y más crítico aún; lo planteado en el estudio de la fuerza laboral de Ciberseguridad de ISC, en el cual se determinó que el déficit ya se encuentra por el orden de tres millones de profesionales: “Una escasez de casi tres millones: este número puede parecer abstracto, pero está teniendo un impacto en el mundo real, en las empresas y en las personas que son responsables de su ciberseguridad. Según la encuesta, el 63% de los encuestados informan que sus organizaciones tienen escasez de personal de TI dedicado a la ciberseguridad. Y casi el 60% dice que sus empresas están en niveles moderados de riesgo extremo de ataques de ciberseguridad debido a esta escasez”. [13]. En Asia se concentra el mayor déficit con 2,17 millones, luego Estados Unidos con cerca de medio millón y Latinoamérica con 0,2 millones. Para el caso de Colombia, el déficit se debe en gran parte a la normatividad que ha venido surgiendo en términos de protección de datos personales, ley de transparencia de la información pública y el CONPES 3854 de 2016, los cuales establecen nuevas obligaciones y condiciones para las organizaciones en torno a la seguridad de información.

Por esta razón las instituciones deben acompañar las iniciativas gubernamentales en torno al área y brindarle al estudiante insumos iniciales para permanecer explorando oportunidades como:

- Becas de formación posgradual: MINTIC, Colciencias – línea de investigación en ciberseguridad y ciberdefensa [14], Fulbright, Colfuturo, etc.
- Becas de cursos, eventos y certificaciones: MinTIC y OEA symposium, Cybersecurity women challenge, OWASP Latam Tour, Foros Universitarios en seguridad Digital (Universidad de los Andes).
- La Universidad de los Llanos brinda matrícula gratuita a cualquiera de sus programas de posgrado al mejor estudiante por facultad, que haya obtenido promedio de carrera igual o superior a 4.5. (Reglamento Estudiantil Universidad de los Llanos – Acuerdo superior 015 de 2003, Artículo 68) [15].
- Cursos en posgrado como opción de grado de los programas de Ingeniería. Electivas de Ciberseguridad ofensiva, criptografía y gestión de seguridad, permiten que el estudiante de grado amplíe su perfil.

III. DESARROLLO DE LA ESTRATEGIA

Teniendo en cuenta que los estudiantes ingresan al programa de grado en edades que oscilan entre los 15 y 17 años de edad y su dedicación es presencial (tiempo completo), la estrategia se ha enfocado en dar continuidad al proceso de formación sobre los aspectos esenciales de ciberseguridad con un enfoque defensivo (protección) a nivel de grado como

Ingeniero de sistemas; para que pueda posteriormente consolidar y continuar su proceso de formación avanzado en otros niveles de formación (Diplomado o posgrado) con un enfoque más ofensivo.

Como parte de este proceso, se diseñó y aplicó un instrumento de levantamiento de información a una muestra del 37% (82 estudiantes de 219) de los estudiantes del programa de Ingeniería de sistemas de la Universidad de los Llanos, una universidad pública de orden nacional, lo que permitió evidenciar aspectos relacionados con el sector de interés para ejercer la profesión y la concienciación sobre las implicaciones de la ciberseguridad en su ámbito profesional, de igual manera permitió brindar a los funcionarios y entes colegiados, responsables del diseño curricular, elementos para la toma de decisiones sobre la adopción de los componentes de ciberseguridad en los programas académicos de grado y posgrado con sus respectivos planes de estudio.

También se revisó un resultado de autoevaluación del plan de estudios del programa de ingeniería de sistemas realizado en el año 2017, donde se relacionaron varias preguntas, que reflejan su grado de variación y permiten evaluar las estrategias tomadas en su momento.



Figura 2. Semillero de Investigación CYSETH

Las acciones han sido tomadas desde un semillero de Investigación denominado CYSETH Unillanos (CyberSecurity & Ethical Hacking Unillanos), conformado por cuatro profesores, 11 estudiantes (IV –VII semestre) y dos egresados, a partir de las siguientes estrategias adoptadas:

TABLA III.
 ESTRATEGIAS Y RESULTADOS PLAN DE CIBERSEGURIDAD
 UNIVERSIDAD DE LOS LLANOS

Estrategia	Resultados
1. Apoyo a la oficina de T.I. de la Universidad (más de mil empleados y siete mil estudiantes). Diseño de políticas, análisis de riesgos.	Construcción de la política de seguridad de información de la Universidad, Apoyo a la implementación de ISO 27005 e ISO/IEC 27001 (en proceso). Profesor auditor certificado ISO27001 y estudiantes.
2. Auditorías a dependencias como proyecto final de un curso en décimo semestre.	Auditorías de controles de seguridad en el marco de ISO/IEC 27001-27002, como proyecto final del curso seguridad de la información en décimo semestre, desde el año 2015 a empresas de la ciudad (más de 50 auditorías en los últimos 4 años).
3. Oferta de cursos en posgrado	Oferta de dos cursos electivos de seguridad ofensiva que se han dictado en cinco cohortes.
4. Diseño curricular en pregrado y posgrado.	Participación en el diseño de 12 cursos que proveen requisitos para asumir cursos de ciberseguridad. Así como dos cursos específicos en pregrado y dos en posgrado.
5. Oferta de diplomados.	Diseño y ejecución de un programa de diplomado “Arquitectura de

	software y seguridad”. 15 estudiantes certificados.
6. Semillero de investigación. (Apoyo de estudiantes antiguos para liderar los nuevos).	Creación de un Semillero denominado CYSETH, conformado por 11 estudiantes y 4 profesores. Vinculación de 15 estudiantes a Proyectos de investigación de GITECX (Grupo de Investigación reconocido por Colciencias).
7. Grupos de interés – capítulo OWASP, rama IEEE, Capítulo ACM.	Búsqueda de apoyo y aval institucional para iniciar la creación del capítulo OWASP regional. Y alianzas con otras iniciativas, que ya existen en la Universidad.
8. Conferencias y demostraciones de Ethical Hacking y blockchain en diferentes semestres.	5 conferencias ofrecidas de forma gratuita (3 internas y dos externas – Universidad Cooperativa y Uniminuto) en ciberseguridad, hacking ético, ataques a la autenticación y blockchain.
9. Eventos, retos e invitados especializados.	Se realizó un foro regional de ciberseguridad, un CTF y apoyo a 4 congresos (CTA – Congreso de tecnologías abiertas) donde se incluye el tema de ciberseguridad (conferencistas incluidos).
10. Encuestas y análisis del contexto.	s Dos instrumentos diseñados y aplicados a más de 130 estudiantes, los cuales fueron aplicados en 2017 y 2019
11. Banco de proyectos para Trabajo de grado.	9 proyectos ejecutados desde el grupo de investigación GITECX y 7 proyectos asignados desde el semillero CYSETH.

Se diseñó un instrumento de levantamiento de información que se aplicó de manera presencial a estudiantes del programa de ingeniería de sistemas, de todos los semestres, buscando indagar tanto por los aspectos propios del diseño curricular y de la perspectiva laboral.

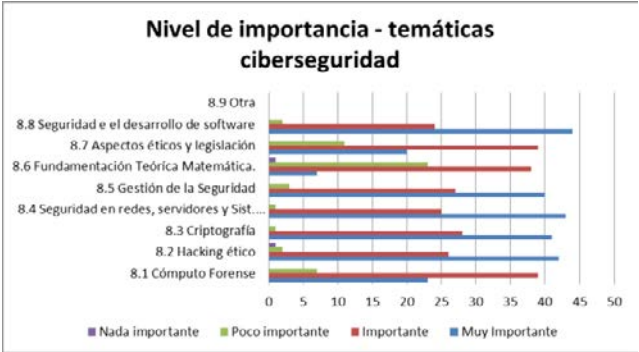


Figura 3. Resultados Preg. Nivel de importancia temática ciberseguridad.

En cuanto a lo curricular, se establecieron varios aspectos relacionados con la caracterización de los estudiantes, así como la importancia y necesidad de los temas en el plan de estudios. Por otro lado, también se indagaron aspectos específicos del agrado por las áreas y sectores del ejercicio profesional.

Muchos estudiantes son proclives al aprendizaje a través de lo praxiológico y manifestaron una preferencia por lo práctico (59%) sobre lo teórico (41%). Esto se puede relacionar con lo indagado sobre la relevancia de las temáticas que debe incluir un curso de ciberseguridad – ver figura 3, encontrando que los

temas asociados directamente con lo tecnológico (Hacking ético, criptografía y Seguridad en redes, servidores y sistemas operativos) son considerados por los estudiantes, como más importantes. Por otro lado, llama la atención que los fundamentos matemáticos, así como los elementos éticos y legales tengan una débil percepción de importancia, dentro de los cursos de seguridad; en el sentido de que la criptografía, catalogada como muy importante; requiere bastante fundamentación matemática y el hacking ético una sólida base ética y legal. De ahí la necesidad de que los profesores generen esta concienciación y orienten el diseño curricular ante estas dualidades.

Teniendo en cuenta que el instrumento aplicó preguntas con respuestas multidimensionales (cada tema con múltiples opciones de nivel de importancia), se calculó un valor total ponderado por cada temática de la siguiente manera:

$$Total\ Ponderado = \sum (Nivel\ Importancia * Factor\ Ponderación).$$

Donde al nivel Muy Importante se le asignó un factor de 1 (100%), al nivel Importante 0,7 (70%) al poco Importante 0,3 (30%) y al nada Importante 0 (0%). Obteniendo los valores de la tabla 4.

TABLA IV.
 RESULTADOS DE PONDERACIÓN TOTAL POR TEMAS.

Temas	Muy Importante	Importante	Poco importante	Nada importante	Total Ponderado
Cómputo Forense	23	39	7	0	60,3
Hacking ético	42	26	2	1	71,2
Criptografía	41	28	1	0	71,2
Seguridad en redes, servidores y Sist. Operat.	43	25	1	0	71,2
Gestión de la Seguridad	40	27	3	0	69,9
Fundamentación Teórica Matemática.	7	38	23	1	48
Aspectos éticos y legislación	20	39	11	0	57,5
Seguridad e el desarrollo de software	44	24	2	0	70,8
Otra	0	0	0	0	0



Figura 4. Estudiantes Décimo semestre, Curso de Seguridad de información, Ingeniería de sistemas.

Ante la diversidad de temas de impacto o emergentes que se deben asumir en los planes de estudio de los programas de T.I., es necesario determinar la percepción e interés, para enfocar los cursos de profundización y electivos en los diferentes niveles de formación. El estudio determinó que el 67% de los estudiantes considera que la “Ciberseguridad y hacking ético” junto a “Big Data” son los temas de mayor impacto y deben ser incluidos en el plan de estudios de manera preferencial, seguidos por “computación inteligente” con 65%, como se ilustra en la figura 5. Las Temáticas asociadas a la reutilización de componentes, educación virtual y Aplicaciones a la agroindustria no superan el 15% de interés.

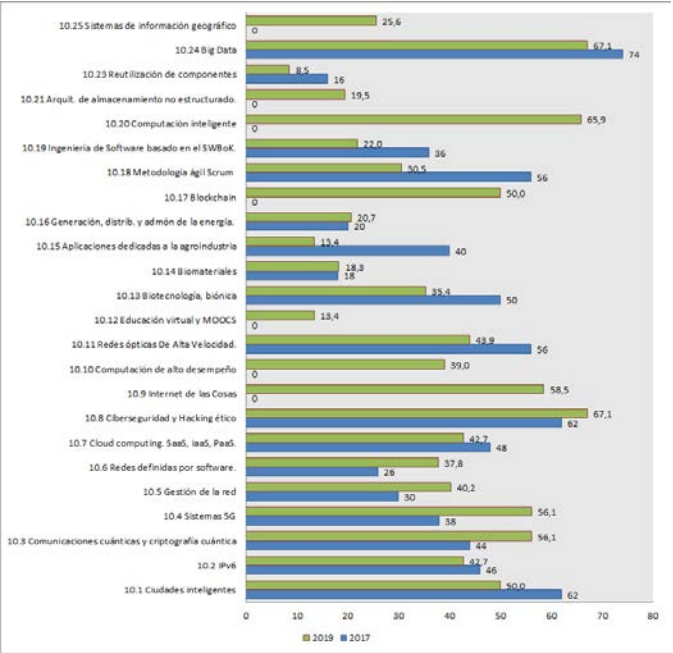


Figura 5.Resultados 2017 y 2019, Temas de impacto que deben incluirse en el plan de estudios

En lo referente contexto de la industria y el sector donde les interesa ejercer la profesión; el instrumento permitió determinar cómo los estudiantes perciben la incidencia de la adopción tecnológica y legislación (que impone ciertas condiciones), y su posible incidencia en la escogencia del sector de interés para ejercer la profesión.

El 74,4% de los estudiantes consideran que el comercio electrónico y las criptomonedas son el principal factor que incidirá en su ejercicio profesional como futuro Ingeniero de sistemas, de igual manera la normatividad del sector TIC (73,2%), Incremento del uso de aplicaciones móviles (63,4%) y el Incremento de los delitos y riesgos cibernéticos (52,4%), según lo indicado en la figura 6.



Figura 6. Resultados Percepción, eventos que direccionarán el ejercicio de la Ingeniería de sistemas.

Para la definición de los sectores indagados a los estudiantes, respecto a la preferencia para ejercer la profesión de ingeniero de sistemas, se tomaron los definidos por el MINTIC en el documento de caracterización de la industria TIC en Colombia [16], por lo cual se asumió el nombre “seguridad informática” literalmente, siendo el más afín a ciberseguridad.

Se consultó finalmente por 28 sectores de la industria T.I., encontrando que los sectores de mayor preferencia son Desarrollo y fabricación de software con el 73,1%, seguido por seguridad informática con el 65,9% y Animación Digital/Video Juegos con el 62,2%. Lo cual es coherente con las iniciativas que el MINTIC viene desarrollando específicamente en estos sectores.

IV. DISCUSIÓN

Los parámetros del contexto, así como los temas relevantes y preferencias estudiantiles en el campo de las TIC, deben tener un seguimiento para identificar tendencias y las motivaciones sobre ellas.

Cuando se plantea formación integral, es esencial incluir en los currículos el componente ético y humanístico, lo cual también está determinado como criterio de acreditación para programas de ciberseguridad en ABET-CSAB (Cybersecurity and Similarly Named Computing Programs: criterio f. Seguridad Humana y criterio h. Seguridad social) [4], que a su vez se encuentran alineados con las guías curriculares de ACM-IEEE CSEC 2017 (Seguridad en el recurso humano – Privacidad y seguridad en el Comportamiento social, Seguridad de Software – Ética, especialmente en el desarrollo, pruebas y revelación de vulnerabilidades).

Para el caso de la Universidad de los Llanos, se creó un nuevo curso en el plan de estudios denominado “ética y humanística” en séptimo semestre y se adopta un capítulo de ética y aspectos legales en todos los cursos relacionados con la ciberseguridad, lo cual evidencia un incremento del 54% al 57% en la cantidad de estudiantes que consideran suficiente el componente “ético y humanístico” en el plan de estudios (Figura 7). Sin embargo se deben desarrollar nuevas estrategias para seguir incrementando este indicador que aún no es el ideal.



Figura 7. Comparación resultados Percepción estudiantes ética y humanidades.

Para el caso de los sectores de la industria de preferencia para ejercer la profesión, se retomó el estudio de autoevaluación de 2017 y se cruzaron los resultados de 2019 sobre los 16 sectores de la industria T.I (MINTIC) que habían sido indagados en ese momento; lo que permitió analizar tendencias y se evidenció (ver figura 8) que **“de los cinco primeros sectores preferidos en el año 2017 (Big data, Ciberseguridad, metodología SCRUM, redes ópticas, ciudades inteligentes); todos descendieron en interés excepto ciberseguridad, que presentó un incremento de 5.1%”**, esto en parte debido a las actividades curriculares y extracurriculares promovidas por los actores interesados (profesores estudiantes y semilleros) en la Universidad.

De los 28 sectores, se incluyeron 12 sectores nuevos en el estudio actual; en los resultados, uno de ellos se ubicó en el primer lugar general (Desarrollo y fabricación de software) y otro en el tercero (animación digital/videojuegos), ambos con más del 60% de aceptación. Mientras que siete de estos sectores nuevos estuvieron por debajo del 20%. Vale la pena resaltar que los sectores desarrollo de software y desarrollo de aplicaciones Web, son dos temas que han dinamizado de manera importante el área de la ciberseguridad a través del proyecto OWASP, que asume de manera directa estos aspectos con enfoque de seguridad y así mismo es una realidad de contexto, que ha presentado un aumento del 56% en los ataques web durante el 2018 [17].

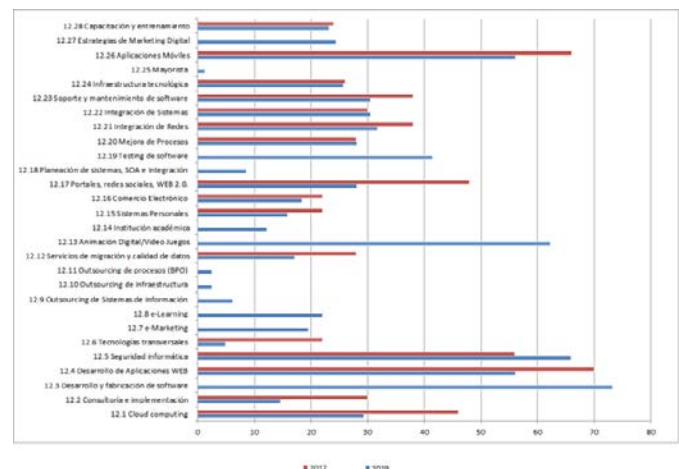


Figura 8. Sectores de la industria T.I. donde los estudiantes prefieren ejercer la profesión (Comparativa 2017 - 2019).

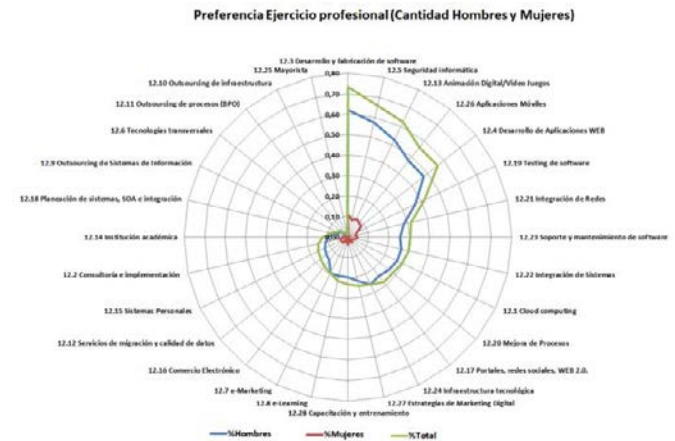
Al analizar los resultados de las preferencias de los sectores para ejercer la profesión y la percepción del grado de orientación que los eventos sociales y de gobierno puedan dar al sector de las TIC en Colombia, se procede a cruzarlas, y tratar de determinar posible incidencia mutua.

Sector Industria Software y TI donde le gustaría ejercer la profesión

Sector	Incremento del uso de aplicaciones móviles	Comercio electrónico y criptomonedas	Incremento del Delitos y riesgos cibernéticos	Total Seguridad Informática
Desarrollo y fabricación de software	10%	70%	10%	80%
Testing de software	10%	30%	10%	30%
Mejora de Procesos	10%	10%	10%	10%
Portales, redes sociales, WEB 2.0.	10%	10%	10%	10%
Capacitación y entrenamiento	10%	10%	10%	10%
Estrategias de Marketing Digital	10%	10%	10%	10%
Aplicaciones Móviles	10%	70%	10%	30%
Infraestructura tecnológica	10%	10%	10%	10%
Soporte y mantenimiento de software	10%	10%	10%	10%
Integración de Redes	10%	10%	10%	10%
Integración de Sistemas	10%	10%	10%	10%
e-Learning	10%	10%	10%	10%
Desarrollo de Aplicaciones WEB	10%	10%	10%	10%
Seguridad Informática	10%	10%	10%	10%
Animación Digital/Video Juegos	10%	10%	10%	10%
Cloud computing	10%	10%	10%	10%

Otro aspecto que se viene promoviendo desde la OEA y ha encontrado eco en los gobiernos y organizaciones civiles, es la necesidad de brindar mayor inclusión a la mujer en el campo de la ciberseguridad. Para lo cual se han aplicado estrategias como el Cyberwomen Challenge en diferentes países (Colombia, Chile, Ecuador, Argentina, Estados Unidos), otorgando becas y recursos financieros. Sin embargo el problema no es solo para el campo de la ciberseguridad, sino de la ingeniería en general; para el caso particular de este proceso, la distribución porcentual de cantidad de estudiantes del programa de ingeniería de sistemas corresponde a 84.9% hombres y 15.06% mujeres y la población tomada para la muestra en este estudio, correspondió a 84% hombres y 16% mujeres.

Sin embargo al observarlo desde el total de la población que escogieron cada sector, la representatividad de mujeres es reducida (ver figura 10): Desarrollo y fabricación de software fue 15%, Animación Digital/Video Juegos con 10%, seguridad informática (9%) y aplicaciones móviles (9%). Según el estudio de ISC [13], las mujeres representan el 24% de la oferta laboral en ciberseguridad a nivel global, de ahí las iniciativas locales (Universitarias) y de la OEA, que muestran que a nivel local se tiene una brecha de 15 puntos porcentuales. .



V. CONCLUSIONES

La realización de proyectos de final de curso o carrera deben estar dirigidos al análisis y solución de problemas del contexto real, la realización de auditorías de seguridad en el marco de una estrategia pedagógica (proyecto de final de curso de profundización en Seguridad de Información en una empresa de la región) es un factor integrador entre la academia y la industria, que proyecta ampliamente al estudiante para continuar laborando o estudiando en el sector.

197

Es importante conformar equipos de trabajo curricular que analicen y construyan todo un proyecto de formación en ciberseguridad, que trasciende un programa académico y un nivel de formación; desplegando competencias, temáticas y componentes curriculares en múltiples cursos de programas de grado, posgrado y no formal (virtual, MOOCS, curso libres). De la misma manera participar activamente en el diseño de los cursos que proveen los requisitos previos (como Sistemas Operativos, Redes, Programación, entre otros).

Promover el interés de los estudiantes involucrándolos en los procesos de construcción curricular con instrumentos de levantamiento de información, invitándolos a cursos y talleres abiertos, demostraciones en vivo y retos CTF (Capture The Flag), entre otros, por fuera de los espacios de clase y presión de la calificación, lo que permite conocer sus intereses particulares y grupales, para tenerlos en cuenta en las estrategias (Caso: Los estudiantes más antiguos, lideran y orientan a los más nuevos).

AGRADECIMIENTOS

Los autores expresan su gratitud a Dios y sus Familias, quienes siempre apoyaron este proceso. De igual manera, a la Universidad de los Llanos, al semillero de investigación CYSETH y al grupo de Investigación en tecnologías abiertas - GITECX, por creer en este Proyecto.

REFERENCIAS

- [1] CISCO Cybersecurity, "Anticipating the Unknowns: CISO Benchmark Study March 2019," 2019.
- [2] Verizon, "Verizon: 2019 Data Breach Investigations Report," 2019.
- [3] R. K. Raj *et al.*, "Perspectives on the future of cybersecurity education," *Proc. - Front. Educ. Conf. FIE*, vol. 2017-Octob, pp. 1–2, 2017.
- [4] ABET, "Computing Accreditation Commission CRITERIA FOR ACCREDITING COMPUTING PROGRAMS 2019-2020," 2019.
- [5] C. S. I.-A. ACM, *Information Technology Curricula 2017: Curriculum Guidelines for Baccalaureate Degree Programs in Information Technology*. 2017.
- [6] A. ACM, IEEE, CSEC - *Curriculum Guidelines for Post-Secondary Degree Programs in Cybersecurity*, vol. Version 1., no. November. 2017.
- [7] Computer Society IEEE - ACM, "CS2013: Computer Science Curricula 2013," *Computer*, vol. 48, no. 3. 2013.
- [8] "PTES Technical Guidelines - The Penetration Testing Execution Standard." [Online]. Available: http://www.pentest-standard.org/index.php/PTES_Technical_Guidelines.

[Accessed: 30-Sep-2019].

- [9] P. Leaders, M. Meucci, and A. Muller, "OWASP, Testing Guide 4.0," no. Cc.
- [10] Departamento Nacional de Planeación, "Conpes 3854 - Política Nacional De Seguridad Digital," 2016.
- [11] ACIS - Asociación Colombiana de Ingenieros de Sistemas, "XIX Encuesta Nacional de Seguridad Informática," *Sistemas*, vol. 151, pp. 12–41, 2019.
- [12] O. Organization of American States, "Ciberseguridad marco NIST, Un abordaje integral de la Ciberseguridad," 2019.
- [13] C. W. Study, "Cybersecurity Professionals Focus on Developing New Skills as Workforce Gap Widens Table of Contents," 2018.
- [14] Colciencias, "Programas Nacionales de CTel." [Online]. Available: <https://www.colciencias.gov.co/investigadores/programas-nacionales-ctei>. [Accessed: 30-Sep-2019].
- [15] Universidad de los Llanos, *Reglamento estudiantil*. 2003, p. 20.
- [16] F. MinTIC, "Caracterización del sector teleinformática, Software y TI en Colombia," 2015.
- [17] Symantec, "ISTR Informe sobre las Amenazas para la Seguridad en Internet," 2019.

Felipe Andrés Corredor Magister en Software libre; del área de administración de redes y sistemas operativos. Especialista en Soluciones telemáticas e Ingeniero de Sistemas. Docente de planta e investigador de la Escuela de ingeniería de la **Universidad de los Llanos**. Sus áreas de desempeño son la seguridad informática y los sistemas distribuidos. Lidera el Grupo de investigación en Tecnologías abiertas GITECX, reconocido por Colciencias.

Diana Cristina Franco Mora Magister en Software libre; del área de administración de redes y sistemas operativos. Especialista en Soluciones telemáticas e Ingeniera de Sistemas. Docente de tiempo completo e investigadora de la Escuela de ingeniería de la **Universidad de los Llanos**. Sus áreas de desempeño son la teleinformática y el Software libre. Es integrante activa del Grupo de investigación en Tecnologías abiertas GITECX, reconocido por Colciencias.

Javier Eduardo Martínez Baquero, Magister en Tecnología educativa; Especialista en Instrumentación Industrial. Docente de planta e investigador de la Escuela de ingeniería de la **Universidad de los Llanos**. Director de la Especialización en Instrumentación y control industrial. Sus áreas de desempeño son la Automatización, Instrumentación. Es integrante activo del Grupo de investigación en Tecnologías abiertas GITECX, reconocido por Colciencias.