RNPS: 0514 ISSN: 2224-6274 VOL.22 - NO. 2 **JULIO / DICIEMBRE 2025**

tono

Revista Científico-Técnica de la Empresa de Telecomunicaciones de Cuba S.A.

ETECSA
EMPRESA DE TELECOMUNICACIONES DE CUBA S.A.

tono

Revista Científico-Técnica

Publicación Semestral Vol.22 – No. 2 - 2025

RNPS: 0514 ISSN: 2224-6274

Tono, Revista Científico-Técnica de la Empresa de Telecomunicaciones de Cuba S.A.

Las opiniones de los autores expresadas en los artículos reflejan sus puntos de vista, pero no necesariamente coinciden con los criterios del consejo editorial.

Los artículos en esta publicación han sido sometidos a revisión por pares a doble ciego.

Portada:
IA

Contáctenos

Dirección de Vigilancia e Información Científico-Técnica de ETECSA
Dirección: Centro de Negocios Miramar, calle 3ra, e/ 76 y 78, Edificio Beijing, 4to Piso, oficina 404. Playa, C.P.: 11300. La Habana, Cuba.
Teléfono : (53) 7266-8453
Correo electrónico: tono@etecsa.cu
Sitio web: www.revistatono.etecsa.cu

CARTA DEL EDITOR

Se ofrece en esta oportunidad una edición especial de la revista *Tono* que compila los principales aportes presentados en el evento de Ciberseguridad en las Telecomunicaciones 2025. Esta selección representa un balance actualizado de los avances y desafíos en esta área prioritaria para el desarrollo tecnológico soberano.

La ciberseguridad se posiciona como elemento transversal en el proceso de digitalización de la sociedad cubana. Los estudios incluidos en esta compilación confirman su carácter determinante para la protección de datos sensibles, la operación segura de servicios fundamentales y la preservación de la integridad del espacio digital nacional. Las investigaciones reflejan respuestas coherentes ante amenazas contemporáneas que exhiben creciente sofisticación.

La práctica de la ciberseguridad en el contexto nacional se distingue por su enfoque sistémico e integrador. Las propuestas documentadas en el evento muestran soluciones que equilibran innovación tecnológica y sostenibilidad práctica, con especial atención a las condiciones específicas del país. Resulta notable el progreso alcanzado en dominios como la seguridad operacional, la gobernanza de datos y la creación de capacidades especializadas.

Los hallazgos del encuentro ratifican la necesidad de mantener un enfoque que combine perfeccionamiento técnico con desarrollo de competencias ciudadanas. Se constata que la eficacia de las estrategias de protección digital requiere de manera simultánea de sistemas tecnológicos avanzados y de usuarios conscientes y preparados.

Esta compilación certifica el nivel de especialización y madurez alcanzado por esta comunidad de conocimiento en Cuba. Se valora el aporte colectivo de investigadores y profesionales cuyos trabajos contribuyen a fortalecer las bases para un desarrollo digital nacional seguro y confiable.

Grupo Editorial
Revista Científico-Técnica Tono

SUMARIO

VOL.22 NO. 2
JULIO / DICIEMBRE 2025

INVESTIGACIÓN

SUMARIO

VOL.22 NO. 2
JULIO / DICIEMBRE 2025

INVESTIGACIÓN

Solución autonómica para la gestión integral de la ciberseguridad basado en aprendizaje por refuerzo

Autonomic solution for cybersecurity integral management based on reinforcement learning

Ing. Ariel Baloira Reyes[*], Dr.C. Walter Baluja García[2], Dr.C. Caridad E. Anías Calderón[3]

Recibido: 11/2025 | Aceptado: 11/2025 | Publicado: 12/2025

Resumen

La ciberseguridad se ha convertido en una carrera contrareloj. Los ciberdelincuentes mejoran sus herramientas cada día y lanzan múltiples ataques en cortos intervalos de tiempo. Este trabajo presenta los componentes funcionales de un sistema autónomo para la gestión integral de la ciberseguridad. Sus objetivos principales son tres: (1) disminuir los tiempos de detección, análisis y respuesta ante amenazas y vulnerabilidades; (2) automatizar los procesos de gestión de ciberseguridad en entidades cubanas; y (3) mantener la infraestructura segura y estable con la mínima intervención humana posible. Para la creación del sistema se utilizaron soluciones de *software* libre

[*] Universidad Tecnológica de La Habana "José Antonio Echeverría", CUJAE. Calle 114 e/ Ciclovía y Rotonda, Marianao, La Habana, Cuba. ariel@tele.cujae.edu.cu

[2] Ministerio de Educación Superior. Calle 23 e/ F y G, Vedado, La Habana, Cuba. walter@mes.gob.cu

[3] Universidad Tecnológica de La Habana "José Antonio Echeverría", CUJAE. Calle 114 e/ Ciclovía y Rotonda, La Habana, Cuba. cacha@tesla.cujae.edu.cu

y código abierto. Para dotar de autonomía al sistema se desplegaron dos modelos de inteligencia artificial empleando Python y PyTorch. El desarrollo de este sistema constituye un paso de avance para Cuba hacia la soberanía tecnológica en el área de la ciberseguridad.

Palabras clave: gestión de ciberseguridad, gestión autonómica, aprendizaje por refuerzo, LLM, ciberseguridad

Abstract

Cybersecurity has become a race against time. Cybercriminals improve their tools every day and perpetuate a large number of attacks in a very short period of time. This work presents the functional components of an autonomous system for integrated cybersecurity management. It seeks to (I) reduce detection, analysis, and response times to threats and vulnerabilities; (II) automate cybersecurity management processes in Cuban entities; and (III) maintain secure and stable the infrastructure with minimal human intervention. Free software and open-source solutions were used to create the system. Two artificial intelligence models using Python and PyTorch were deployed to make the system autonomous. The development of this system represents a step forward for Cuba toward technological sovereignty in the area of cybersecurity.

Keywords: *cybersecurity management; autonomous management; reinforcement learning; LLM; cybersecurity*

Introducción

La ciberseguridad actual se ha convertido en una carrera contrarreloj entre los atacantes y los operadores de las infraestructuras y sistemas de TI (Tecnologías de la Información). Las amenazas han aumentado de manera exponencial con el pasar de los años y su control representa un desgaste continuo para los especialistas de ciberseguridad (2024 SonicWall Cyber Threat Report, 2024; CrowdStrike, 2024; European Union Agency for Cybersecurity., 2024; Fabio Assolini, 2024). Los ciberatacantes han escalado sus operaciones por medio de la automatización y el empleo de la inteligencia artificial (IA). Esto obliga a los especialistas de ciberseguridad a desarrollar soluciones automatizadas que permitan hacerle frente.

La investigación realizada por el Centro de Estudios Estratégicos e Internacionales (CSIS) de los Estados Unidos muestra que las grandes organizaciones emplean aproximadamente cuarenta y siete herramientas de ciberseguridad diferentes en sus redes de un promedio de diez proveedores (Crumpler & Lewis, 2020). Gestionar este cúmulo de herramientas conlleva cuantiosos gastos monetarios y humanos, sobre todo para brindar respuesta a las amenazas de forma proactiva y veloz. Como solución a esta problemática, se ha planteado la integración de diferentes tecnologías como los sistemas de Gestión de Información y Eventos de Seguridad (SIEM, siglas del término en inglés, *Security Information and Event Management*), de Detección y Respuesta Extendidas (XDR, *eXtended Detection and Response*), IA y la computación en la nube (European Union Agency for Cybersecurity., 2023; Uzoma et al., 2023).

La detección oportuna y la identificación precisa de cambios en los patrones de comportamiento en la infraestructura son cruciales para garantizar el funcionamiento efectivo de los mecanismos de protección de ciberseguridad. La naturaleza no estacionaria del ciberespacio plantea un desafío significativo para el desarrollo de soluciones de protección robustas y actualizadas. Por lo tanto, los métodos de IA a emplear necesitan ajustarse de forma dinámica a las condiciones cambiantes del entorno, con el fin de que los sistemas de defensa respondan con eficacia a las amenazas. Como solución se propone el empleo de algoritmos de Aprendizaje por Refuerzo, *Reinforcement Learning*), capaces de generar modelos que imitan el razonamiento humano. Estos son adaptativos, aprenden del entorno en que se encuentran y seleccionan las acciones que maximizan una recompensa acumulada, basándose en la experiencia previa. (Adawadkar & Kulkarni, 2022; Alonso et al., 2021; Hore et al., 2022; Huang et al., 2022; Mathew, 2021; Phan & Bauschert, 2022; Sewak et al., 2022).

Este trabajo presenta un sistema basado en herramientas de *software* libre y código abierto para gestionar información de ciberseguridad, empleando técnicas de RL para responder de forma autónoma a las amenazas existentes en la infraestructura gestionada. Permite a los

especialistas de ciberseguridad supervisar la toma de decisiones del sistema y obtener una respuesta rápida y efectiva a los posibles ciberataques. El sistema diseñado busca alcanzar los siguientes objetivos: (1) disminuir tiempos de detección, análisis y respuesta a amenazas y vulnerabilidades; (2) automatizar procesos de gestión de ciberseguridad en entidades de Cuba; y (3) mantener la infraestructura segura y estable con mínima intervención humana.

Materiales y métodos

En la presente investigación, se emplearon los métodos de investigación científica de análisis-síntesis para el estudio de los componentes de soluciones SIEM y XDR necesarios para implementar la gestión automatizada de la ciberseguridad en entornos heterogéneos. Se recurre al método del modelado y simulación computacional para representar los ataques cibernéticos que el sistema tiene que detectar y mitigar de forma automatizada. Por último, se emplean los métodos empíricos de observación, medición y experimentación durante la validación de la solución propuesta.

Para conformar el sistema se utilizaron las soluciones de *software* libre y código abierto Wazuh, OpenVAS, Suricata y Snort. Para el desarrollo de la solución de IA se usó el lenguaje de programación Python junto con la librería PyTorch como *framework* principal para la construcción de los modelos. La solución propuesta se encuentra en la fase de desarrollo y optimización, no obstante, existe una primera versión de la plataforma operativa.

Resultados y discusión

El sistema para la gestión autónoma de la ciberseguridad está integrado por seis componentes: (1) recolección de datos, (2) monitorización y análisis, (3) detección y clasificación, (4) autonomía decisional, (5) supervisión especializada y (6) almacenamiento. Para garantizar la estabilidad y disponibilidad, su diseño se sustenta en la arquitectura de microservicios. Esto permite la progresión de los componentes y posibilita la recolección de los datos por grupos de colectores de manera simultánea. La interrelación entre componentes se muestra en la Figura 1.

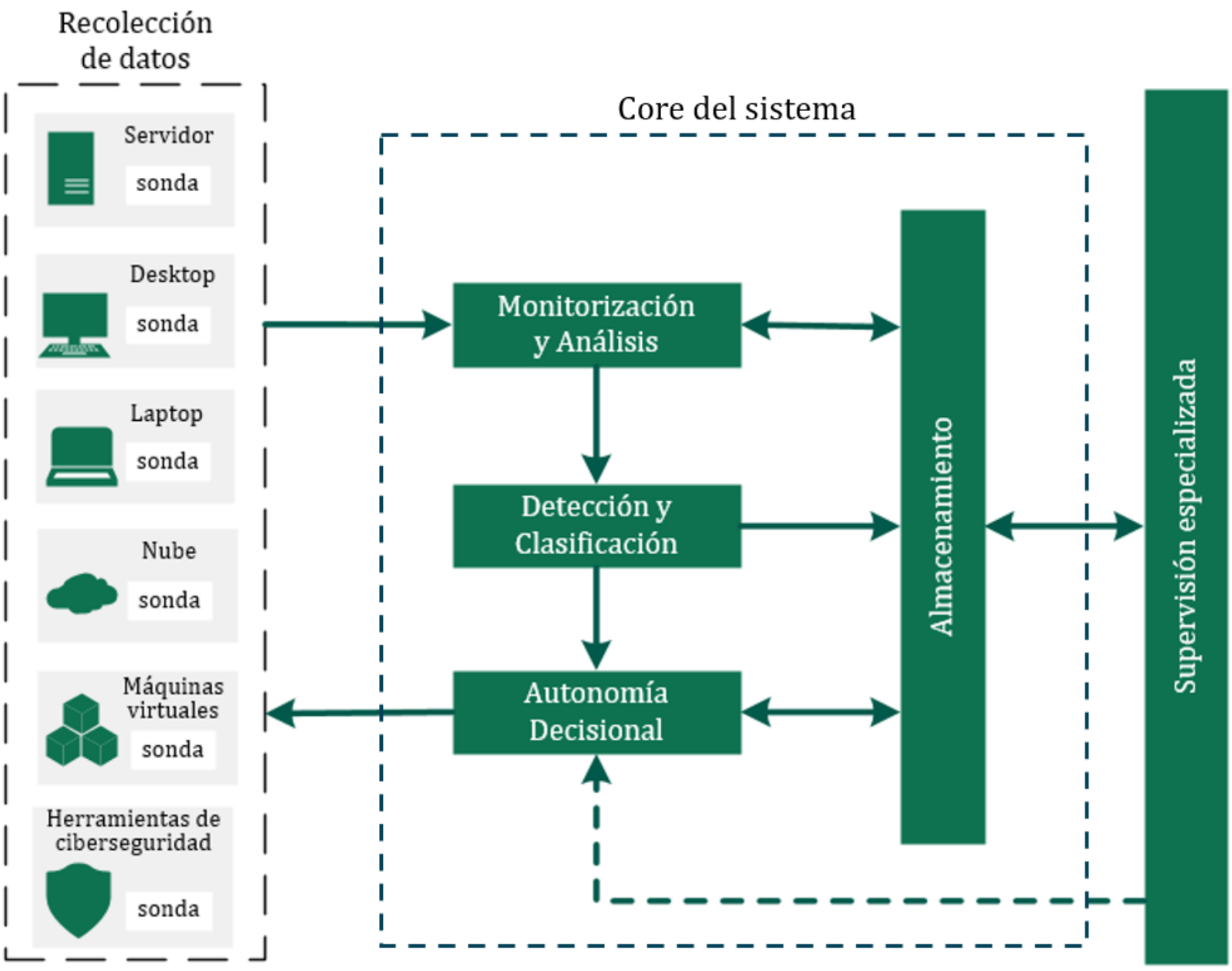

Figura 1. Componentes internos del sistema autónomo de gestión integral de ciberseguridad

Recolección de datos

El componente principal es el de recolección de datos, encargado de la obtención de la información de interés. Está compuesto por sondas distribuidas en la infraestructura de red y agentes de recolección que se encuentran desplegados en los puntos finales, denominados colectores. Su característica fundamental es la capacidad de recopilar datos heterogéneos que incluyen: registros del sistema operativo, tráfico de red, registros de eventos de instancias de virtualización y computación en la niebla/nube, alertas de sistemas de seguridad como los Sistemas de Detección de Intrusos (IDS, *Intrusion Detection System*) o cortafuegos.

Los datos recopilados se pueden clasificar en dos grupos: generales y específicos. Los datos generales están constituidos por registros de tráfico, registros del sistema operativo, registros de aplicaciones y registros de seguridad. Los datos específicos consisten en confirmaciones de existencia de vulnerabilidades, datos de amenazas y datos de inteligencia colaborativa. Es importante recalcar que el tipo específico de datos variará en función de las herramientas de ciberseguridad

que se utilicen y de las necesidades de la organización; sin embargo, los datos generales y específicos descritos anteriormente se consideran esenciales para prevenir y detectar incidentes de ciberseguridad.

En concreto, el objetivo del componente de Recolección de Datos es obtener datos de ciberseguridad en bruto de diversas fuentes, en tiempo real o con un retraso mínimo. Estos serán enviados al componente Monitorización y Análisis para su procesamiento y conservación.

Monitorización y Análisis

Para procesar y almacenar los datos recogidos por los colectores es necesario poseer un punto de intercambio común para el sistema. En este sentido, el componente de Monitorización y Análisis provee interfaces de comunicación estándares con el componente de Recolección de Datos. Estas son *Syslog* (*System Logging Protocol*) que permite el intercambio de mensajes de registro con un formato estándar; y API REST (API, Interfaz de Programación de Aplicaciones basada en la arquitectura de Transferencia de Estado Representacional, REST).

Varios colectores pueden recopilar la misma información y representarla de forma distinta. Por este motivo, una de sus funciones fundamentales es la normalización de la información recibida. Esta función se realiza sustentada en expresiones regulares (Zheng et al., 2021), y en el formato JSON (del término en inglés, *JavaScript Object Notation*); formato de intercambio de datos ligero fácil de leer y escribir para los humanos y las máquinas (Banhara et al., 2023). Con su implementación se puede generar una estructura de datos ordenados, de fácil entendimiento, además de generalizar la información relevante para los procesos de análisis; basado en este primer paso de procesamiento de los datos en bruto, se pueden enriquecer los registros con información asociada. Por ejemplo, la dirección IP de un paquete puede dar a conocer el país y el proveedor de servicios correspondiente.

Otra función importante del componente de Monitorización y Análisis es la correlación de los datos obtenidos de los colectores, para agrupar la información y evitar su duplicación. Dado que las infraestructuras de red pueden variar en tamaños, es necesario que este componente sea capaz de recibir y normalizar un volumen elevado de información de forma simultánea.

Detección y Clasificación

Es el responsable de detectar amenazas, anomalías e infracciones de cumplimiento normativo y de generar alertas al identificar actividad sospechosa. Dicho componente se apoya en múltiples fuentes de inteligencia de amenazas y enriquece las alertas con datos contextuales para mejorar la precisión de la detección. Esto incluye la asignación de eventos al marco MITRE ATT&CK, la detección de vulnerabilidades con el servicio Wazuh CTI y la alineación de los hallazgos con estándares regulatorios como PCI DSS, RGPD, HIPAA, los puntos de referencia CIS y NIST 800-53 (Figura 2). Estas capacidades proporcionan información práctica para la búsqueda de amenazas, la detección de vulnerabilidades y la monitorización del cumplimiento normativo.

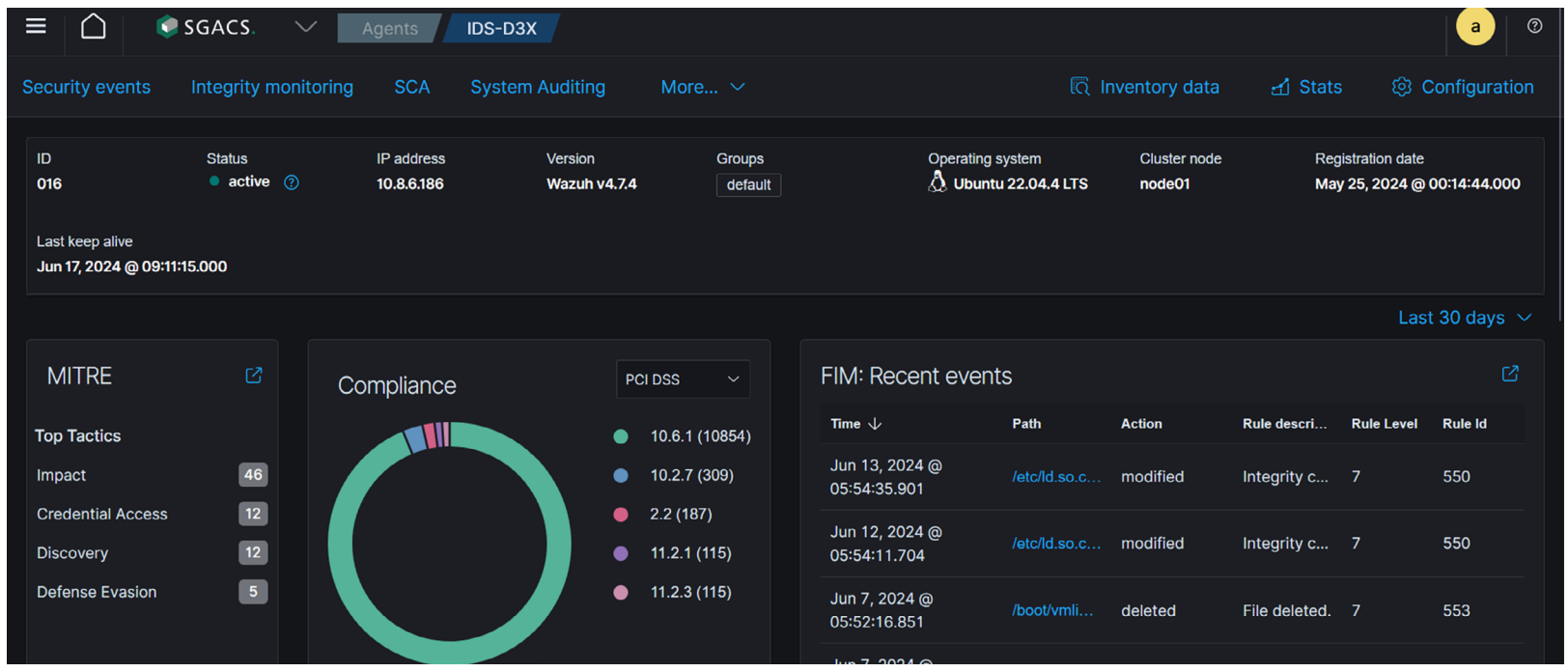

Figura 2. Ilustración del funcionamiento del componente de Detección y Clasificación

Mediante los complementos de la plataforma Wazuh, el componente de Detección y Clasificación se integra con plataformas externas para optimizar los flujos de trabajo. Algunos ejemplos incluyen sistemas de *tickets* como Jira o TheHive, así como herramientas de comunicación como Slack o Telegram.

Un valor muy importante lo tienen los modelos de Aprendizaje Automático (ML, siglas del término en inglés, *MachineLearning*). Estos se implementan como piezas clave en la detección y clasificación de amenazas. Para la implementación de este componente, se seleccionó el algoritmo kNN (*k-nearest neighbors*) que es un clasificador de aprendizaje supervisado que emplea la proximidad para realizar clasificaciones o predicciones sobre la agrupación de un punto de datos individual.

Autonomía decisional

El cerebro del sistema está representado por el componente Autonomía Decisional. Este se forma por dos tecnologías de IA: (1) Modelos Grandes de Lenguaje (LLM, *Large Language Models*) y (2) Aprendizaje por Refuerzo Profundo (DRL, *Deep Reinforcement Learning*). El LLM recibe información del componente Detección y Clasificación, la procesa y organiza en una matriz de observación que funciona como entrada del modelo DRL. Este último decide qué acción tomar para disminuir o mitigar los riesgos. La integración de estos modelos contribuye directamente a la autonomía del sistema.

La matriz de observación representa el estado de la infraestructura gestionada. En consecuencia, el LLM se entrena y prepara para conformar una imagen de la situación actual de la infraestructura. Para conseguirlo, se apoya en eventos pasados, datos del marco MITRE ATT&CK, datos sobre las vulnerabilidades presentes y estado del cumplimiento de las normativas internacionales y nacionales. Dado que el sistema está destinado principalmente para las instituciones cubanas, el LLM ha sido refinado con las regulaciones nacionales referidas a la ciberseguridad. Hasta el momento se ha experimentado con el modelo Llama 3.2:3b. En la actualidad, se ejecuta un estudio de contraste con otros modelos dentro del estado de la cuestión con el objetivo de seleccionar el que mejor se ajuste a los recursos disponibles.

Por otro lado, los estudios realizados por Adawadkar y Kulkarni (2022); Asmat et al. (2025); Hore et al. (2022), Huang et al. (2022) y; Phan y Bauschert (2022) señalan que un modelo de DRL posee la habilidad de tomar decisiones estratégicas y se basa tanto en el estado actual como en las condiciones anteriores del entorno. Para alcanzar este nivel de autonomía, se opta por utilizar una red neuronal de tipo DDQN (*Double Deep Q Network*), cuya principal ventaja radica en su capacidad de mejora continua y que opera como un como un programa que evoluciona de manera similar al aprendizaje humano. En cada iteración, el algoritmo analiza las acciones realizadas en función del entorno observado y ajusta su comportamiento con el objetivo de

maximizar la recompensa obtenida. En la Figura 3 se representa el diagrama de funcionamiento de este componente.

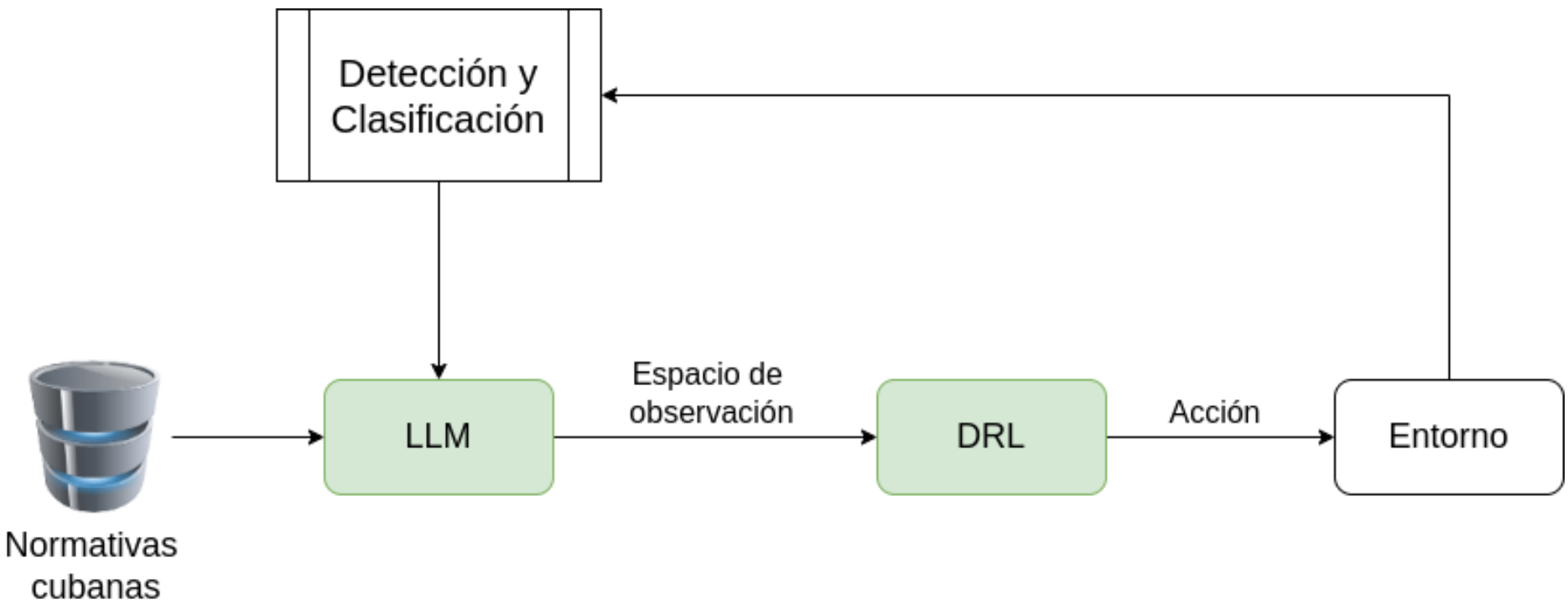

Figura 3. Diagrama de funcionamiento del componente de Autonomía Decisional

Almacenamiento

El componente de Almacenamiento gestiona toda la información generada por los componentes anteriores, utilizando formato JSON para facilitar su indexación. Esta función se implementa mediante OpenSearch (https://opensearch.org/), plataforma de búsqueda y observación de código abierto. La función principal del componente de Almacenamiento es la conservación de los datos, el cumplimiento normativo y el análisis forense. Al ser un sistema distribuido y desplegado sobre la tecnología Docker, el componente ofrece escalabilidad y persistencia. Gracias a la conservación de los datos durante largos períodos de tiempo permite, tanto a especialistas como investigadores, acceder a datos históricos y realizar un análisis forense. Además, posibilita analizar tendencias, identificar patrones recurrentes y evaluar la efectividad de las medidas implementadas. Al mismo tiempo, se garantiza uno de los requisitos de ciberseguridad primordial: la conservación de los datos de eventos para su análisis en situaciones que lo ameriten (*ISO/IEC 27001:2022*, 2022).

Supervisión especializada

Al igual que un docente supervisa el proceso de aprendizaje de sus estudiantes, es necesario implementar mecanismos de control que permitan evaluar las decisiones del sistema en tiempo real y

retrospectivamente. A pesar de los beneficios de la introducción de la IA en la ciberseguridad, la dificultad para entender cómo actúan estos modelos mientras operan de forma autónoma es un riesgo que no puede ser ignorado por los especialistas ni las organizaciones (Arreche et al., 2024; Atakishiyev et al., 2024; Biswas et al., 2024; Nwakanma et al., 2023; Rjoub et al., 2023; Tsakalakis et al., 2024). El componente de Supervisión Especializada es el que permite visualizar las acciones realizadas por el sistema, así como acceder a los datos normalizados y procesados representándolos de forma clara mediante diagramas, gráficos y tablas (Figura 4).

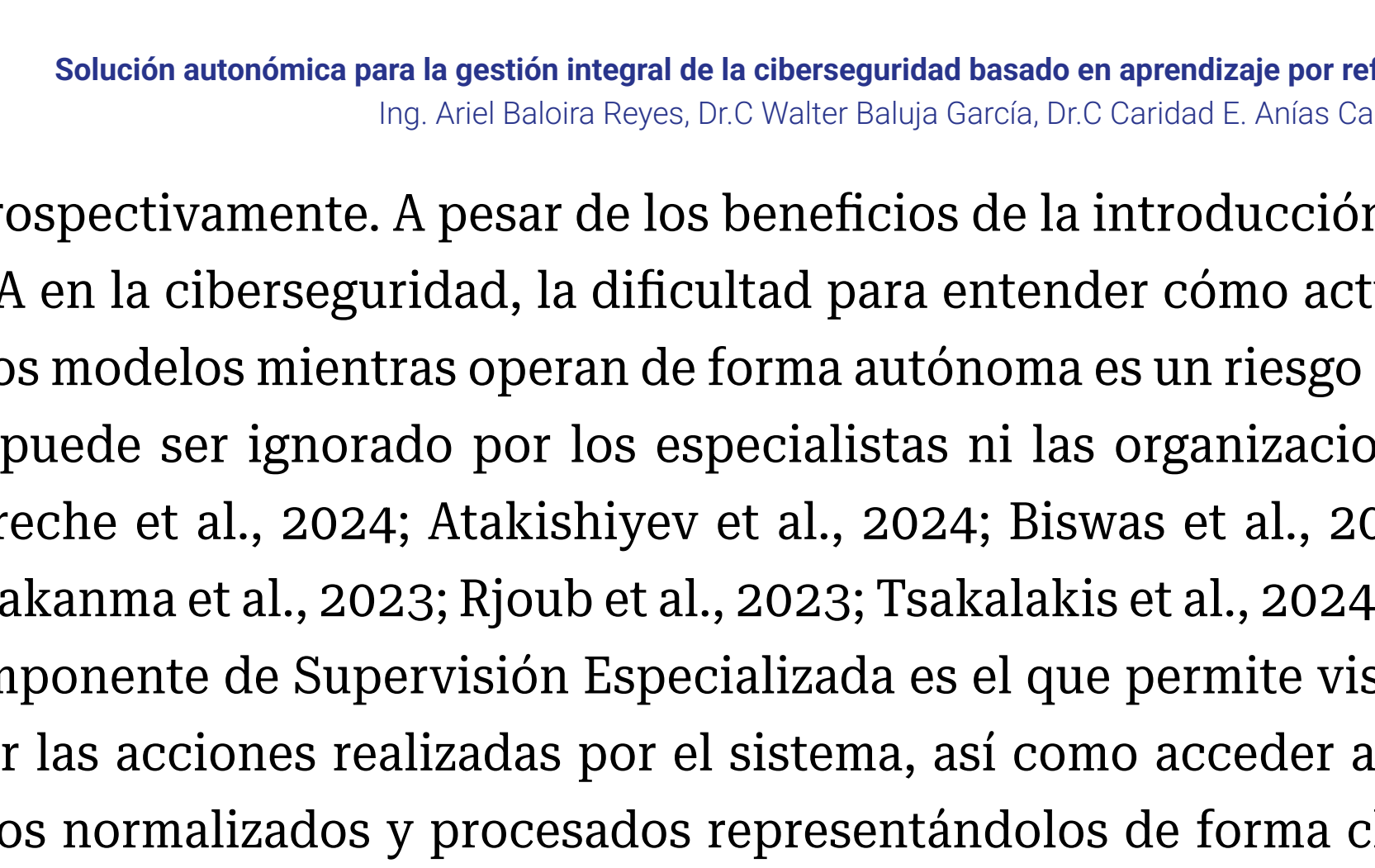

Figura 4. Visualización de las vulnerabilidades presentes en un dispositivo

Es el responsable de crear una representación gráfica intuitiva de la recopilación, el almacenamiento y el procesamiento de los datos de ciberseguridad. Esta representación puede ser aprovechada por las partes interesadas (como analistas de seguridad, gerentes de TI y ejecutivos), para comprender con mayor claridad el panorama de ciberseguridad de la organización y tomar decisiones informadas sobre la gestión de riesgos (Figura 5). Este componente facilita la interpretación de los elementos que desencadenan una acción específica en los modelos de IA, sobre todo el de RL. Su función es primordial para evitar el sobreajuste y corregir desviaciones del sistema de forma anticipada.

›	Jun 28, 2023 @ 21:40:25.861	PAM: User login failed.	5	5503
›	Jun 28, 2023 @ 21:40:25.847	Host Blocked by firewall-drop Active Response	3	651
›	Jun 28, 2023 @ 21:40:25.610	sshd: Attempt to login using a non-existent user	5	5710
›	Jun 28, 2023 @ 21:40:25.607	sshd: Attempt to login using a non-existent user	5	5710
›	Jun 28, 2023 @ 21:40:25.607	sshd: brute force trying to get access to the system. Non existent user.	10	5712
›	Jun 28, 2023 @ 21:40:24.156	sshd: Attempt to login using a non-existent user	5	5710
›	Jun 28, 2023 @ 21:40:24.156	sshd: Attempt to login using a non-existent user	5	5710
›	Jun 28, 2023 @ 21:40:21.847	PAM: User login failed.	5	5503
›	Jun 28, 2023 @ 21:40:21.843	PAM: User login failed.	5	5503
›	Jun 28, 2023 @ 21:40:21.618	sshd: Attempt to login using a non-existent user	5	5710
›	Jun 28, 2023 @ 21:40:21.618	sshd: Attempt to login using a non-existent user	5	5710
›	Jun 28, 2023 @ 21:40:21.618	sshd: Attempt to login using a non-existent user	5	5710

Figura 5. Registros de acciones y eventos del sistema

Conclusiones

El desarrollo de un sistema autónomo para la gestión integral de la ciberseguridad es un paso más hacia la soberanía tecnológica nacional. La integración de técnicas de IA permite reducir los tiempos de respuesta a incidentes y la corrección de vulnerabilidades antes de que sean explotadas. Los seis componentes que conforman el sistema cumplen funciones independientes, sin embargo al integrarse aportan una capacidad significativa a la solución desarrollada. El componente de Autonomía Decisional representa una de las principales contribuciones para lograr la ciberseguridad autónoma, pero implica un riesgo considerable si no se gestiona de forma adecuada. Esta función de control se delega en el componente de Supervisión Especializada, el cual permite la visualización de todas las acciones que realiza el sistema y su supervisión controlada. Es necesario mantener el constante perfeccionamiento de los modelos de lenguaje empleados, con el fin de optimizar el uso de recursos. Como línea de trabajo futura resulta esencial profundizar en el estudio y diseño de modelos de explicabilidad orientados a interpretar las decisiones del sistema y sus implicaciones.

Referencias bibliográficas

2024 SonicWall Cyber Threat Report. (2024). https://www.sonicwall.com/resources/white-papers/2024-sonicwall-cyber-threat-report

Adawadkar, A. M. K., & Kulkarni, N. (2022). Cyber-security and reinforcement learning—*A brief survey*. *Engineering Applications of Artificial Intelligence*, 114, 105116. https://doi.org/10.1016/j.engappai.2022.105116

Alonso, R. S., Prieto, J., La Prieta, F. D., Rodriguez-Gonzalez, S., & Corchado, J. M. (2021). *A Review on Deep Reinforcement Learning for the management of SDN and NFV in Edge-IoT*. 2021 IEEE Globecom Workshops (GC Wkshps), 1-6. https://doi.org/10.1109/GCWkshps52748.2021.9682179

Arreche, O., Guntur, T., & Abdallah, M. (2024). XAI-IDS: Toward Proposing an Explainable Artificial Intelligence Framework for Enhancing Network Intrusion Detection Systems. *Applied Sciences, 14(10), 4170*. https://doi.org/10.3390/app14104170

Asmat, H., Din, I. U., Almogren, A., & Khan, M. Y. (2025). *Digital Twin with Soft Actor-Critic Reinforcement Learning for Transitioning from Industry 4.0 to 5.0*. IEEE Access, 1-1. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3546085

Atakishiyev, S., Salameh, M., Yao, H., & Goebel, R. (2024). *Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions*. IEEE Access, 12, 101603-101625. https://doi.org/10.1109/ACCESS.2024.3431437

Banhara, N., Duarte, D., & Schreiner, G. (2023). Extração de Esquemas de Documentos JSON: O que há de Novo? *Escola Regional de Banco de Dados (ERBD)*, 11-20. https://doi.org/10.5753/erbd.2023.229421

Biswas, B., Mukhopadhyay, A., Kumar, A., & Delen, D. (2024). *A hybrid framework using explainable AI (XAI) in cyber-risk management for defence and recovery against phishing attacks*. Decision Support Systems, 177, 114102. https://doi.org/10.1016/j.dss.2023.114102

CrowdStrike. (2024). *CrowdStrike 2024 Global Threat Report*. https://www.crowdstrike.com/en-us/global-threat-report/

Crumpler, W. D., & Lewis, J. A. (2020). *Cybersecurity and the Problem of Interoperability*.

European Union Agency for Cybersecurity. (2023). Artificial intelligence and cybersecurity research: ENISA research and innovation Brief. *Publications Office*. https://data.europa.eu/doi/10.2824/808362

European Union Agency for Cybersecurity. (2024). ENISA threat landscape 2024: July 2023 to June 2024. *Publications Office*. https://data.europa.eu/doi/10.2824/0710888

Fabio Assolini. (2024, marzo 19). *Panorama de Amenazas Cibernéticas en Cuba y región* [Conferencia magistral]. https://www.informaticahabana.cu/actividad/panorama-de-amenazas-ciberneticas-en-cuba-y-region/

Hore, S., Shah, A., & Bastian, N. D. (2022). *Deep VULMAN: A Deep Reinforcement Learning-Enabled Cyber Vulnerability Management Framework* (No. arXiv:2208.02369). arXiv. https://doi.org/10.48550/arXiv.2208.02369

Huang, Y., Huang, L., & Zhu, Q. (2022). Reinforcement Learning for feedback-enabled cyber resilience. *Annual Reviews in Control*, 53, 273-295. https://doi.org/10.1016/j.arcontrol.2022.01.001

ISO/IEC 27001:2022—Information security management systems. (2022). [International Standard published]. ISO/IEC JTC 1/SC 27. https://www-iso-org.translate.goog/standard/27001?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc

Mathew, A. (2021). Deep Reinforcement Learning for Cybersecurity Applications. *International Journal of Computer Science and Mobile Computing*, 10(12), 32-38. https://doi.org/10.47760/ijcsmc.2021.v10i12.005

Nwakanma, C. I., Ahakonye, L. A. C., Njoku, J. N., Odirichukwu, J. C., Okolie, S. A., Uzondu, C., Ndubuisi Nweke, C. C., & Kim, D.-S. (2023). Explainable Artificial Intelligence (XAI) for Intrusion Detection and Mitigation in Intelligent Connected Vehicles: A Review. *Applied Sciences*, 13(3), 1252. https://doi.org/10.3390/app13031252

Phan, T. V., & Bauschert, T. (2022). DeepAir: Deep Reinforcement Learning for Adaptive Intrusion Response in Software-Defined Networks. *IEEE Transactions on Network and Service Management*, 19(3), 2207-2218. https://doi.org/10.1109/TNSM.2022.3158468

Rjoub, G., Bentahar, J., Wahab, O. A., Mizouni, R., Song, A., Cohen, R., Otrok, H., & Mourad, A. (2023). A Survey on Explainable Artificial Intelligence for Cybersecurity. *IEEE Transactions on Network and Service Management*, 20(4), 5115-5140. https://doi.org/10.1109/TNSM.2023.3282740

Sewak, M., Sahay, S. K., & Rathore, H. (2022). Deep Reinforcement Learning for Cybersecurity Threat Detection and Protection: A Review. En R. Krishnan, H. R. Rao, S. K. Sahay, S. Samtani, & Z. Zhao (Eds.), *Secure Knowledge Management In The Artificial Intelligence Era* (Vol. 1549, pp. 51-72). Springer International Publishing. https://doi.org/10.1007/978-3-030-97532-6_4

Tsakalakis, N., Stalla-Bourdillon, S., Huynh, T. D., & Moreau, L. (2024). *A taxonomy of explanations to support Explainability-by-Design* (No. arXiv:2206.04438). arXiv. https://doi.org/10.48550/arXiv.2206.04438

Uzoma, J., Falana, O., Obunadike, C., Oloyede, K., & Obunadike, E. (2023). *Using artificial intelligence for automated incidence response in cybersecurity*.

Zheng, L.-X., Ma, S., Chen, Z.-X., & Luo, X.-Y. (2021). Ensuring the Correctness of Regular Expressions: A Review. *International Journal of Automation and Computing*, 18(4), 521-535. https://doi.org/10.1007/s11633-021-1301-4

Métodos de autenticación basados en criptografía de conocimiento cero

Zero-knowledge proof-based cryptographic authentication methods

Ing. Adiane Cueto Portuondo[*], Ing. Angel Alejandro Guerra Vilches[2], Ing. Leanny Laura Duardo Polo[3]

Recibido: 11/2025 | Aceptado: 11/2025 | Publicado:12/2025

Resumen

El presente artículo examina la evolución de los métodos de autenticación en un contexto donde la seguridad y la privacidad son cada vez más cruciales. Se destaca la necesidad de sistemas que garanticen la confidencialidad y la integridad de la información, sobre todo en un medio digital vulnerable. El estudio se centra en explorar las distintas categorías de autenticación desde métodos tradicionales, como contraseñas y tókenes físicos, hasta enfoques más avanzados, como la autenticación multifactorial (MFA, *Multi-Factor Autentication*) y el uso de certificados digitales. Sin embargo, el enfoque principal radica en el análisis de las Pruebas de Conocimiento Cero (ZKP, *Zero Knowledge Proof*) como una solución innovadora para mitigar las vulnerabilidades inherentes a los métodos tradicionales. La metodología utilizada incluye una revisión exhaustiva de la literatura y un análisis

[*] Departamento de Ciberseguridad, Facultad Ciberseguridad, Universidad de las Ciencias Informáticas. adianecp855@gmail.com

[2] Departamento de Infraestructuras Tecnológicas, Facultad de Ciberseguridad, Universidad de las Ciencias Informáticas. angelagv@uci.cu

[3] Departamento de Infraestructuras Tecnológicas, Facultad de Ciberseguridad, Universidad de las Ciencias Informáticas. leannyldp@uci.cu

detallado de los protocolos de ZKP aplicados a la autenticación, con especial énfasis en su implementación en entornos de computación en la nube. Los resultados subrayan la eficacia de las ZKP para verificar la identidad de los usuarios sin comprometer la información sensible y presentan ventajas significativas en comparación con otros métodos.

Palabras clave: autenticación, criptografía de conocimiento cero, computación en la nube, seguridad, privacidad

Abstract

The article "Authentication Methods Based on Zero-Knowledge Cryptography" examines the evolution of authentication methods in a context where security and privacy are increasingly crucial. It highlights the need for systems that ensure the confidentiality and integrity of information, especially in an increasingly vulnerable digital environment. The estudy focuses on exploring various categories of authentication, from traditional methods such as passwords and physical tokens to more advanced approaches like Multi-Factor Authentication (MFA) and the use of digital certificates. However, the main focus is on the analysis of Zero-Knowledge Proofs (ZKP) as an innovative solution to mitigate the inherent vulnerabilities of traditional methods. The methodology used includes an exhaustive literature review and a detailed analysis of ZKP protocols applied to authentication, with particular emphasis on their implementation in cloud computing environments. The results underscore the effectiveness of ZKPs in verifying user identities without compromising sensitive information, presenting significant advantages over other methods.

Keywords: *authentication, zero-knowledge cryptography, cloud computing, security, privacy*

Introducción

Un ciudadano promedio utiliza las tecnologías para acceder a todo tipo de servicios (correo, cuentas bancarias, redes sociales, dispositivos móviles, servicios de comunicaciones, entre otros). Estos servicios deben proporcionar seguridad y confiabilidad, para lo cual se requiere

algún tipo de información que identifique de forma única al usuario (Rodríguez Valdés et al., 2018).

La autenticación es un proceso que verifica la identidad de un usuario que intenta acceder a una información específica. Los sistemas de autenticación se utilizan de forma habitual para proteger datos sensibles e impedir el acceso no autorizado a sistemas y redes (Marmolejo-Corona et al., 2023). Este proceso es fundamental para garantizar la confidencialidad, integridad y disponibilidad de los recursos en un entorno digital, pues la autenticación efectiva minimiza el riesgo de accesos no autorizados y protege la información contra robos, alteraciones o usos indebidos. Con el aumento de la ciberdelincuencia y de la exposición de información personal, la necesidad de métodos de autenticación robustos y seguros ha cobrado una importancia crucial. Las técnicas de autenticación han evolucionado desde las contraseñas tradicionales hasta las Pruebas de Conocimiento Cero (ZKP) para enfrentar los desafíos de seguridad en un mundo cada vez más digitalizado.

Este trabajo explora diversos métodos de autenticación, desde los más tradicionales hasta los más innovadores, con un enfoque especial en la criptografía de conocimiento cero. El problema central que aborda la investigación es la vulnerabilidad de los métodos de autenticación tradicionales (como las contraseñas y los tókenes físicos) frente a ataques sofisticados y la exposición de información sensible. El artículo se divide en varias secciones. En primer lugar, se realiza una revisión exhaustiva de la literatura sobre métodos de autenticación, en la que se destacan las limitaciones y vulnerabilidades de los enfoques tradicionales. En segundo lugar, se profundiza en las ZKP, mediante la explicación sus fundamentos teóricos y su aplicación en la autenticación. Se analizan diferentes protocolos de ZKP (como el de Schnorr, Feige-Fiat- Shamir, y Guillou-Quisquater), y se discute su seguridad y eficiencia.

El objetivo final es proporcionar una comprensión profunda de cómo las ZKP pueden utilizarse para asegurar la identidad de los usuarios sin comprometer la privacidad ni la seguridad de la información. Este enfoque innovador no solo mitiga las vulnerabilidades

de los métodos tradicionales, sino que también abre nuevas posibilidades para la autenticación segura en un entorno digital cada vez más complejo y expuesto a amenazas.

Materiales y métodos

Para la elaboración de este estudio se realizó una búsqueda exhaustiva en diversas bases de datos académicas reconocidas como Google Scholar, Scopus, IEEE Xplore y ACM Digital Library.

Con el objetivo de garantizar la inclusión de estudios relevantes y de alta calidad se establecieron criterios de inclusión específicos.

- Publicaciones en revistas revisadas por pares o conferencias de alto impacto: se priorizaron estudios publicados en revistas científicas con un riguroso proceso de revisión por pares, así como en conferencias internacionalmente reconocidas en los campos de la criptografía, la seguridad en la nube y la privacidad de datos.
- Enfoque en métodos de autenticación basados en ZPK: se seleccionaron estudios que abordaran el desarrollo, la implementación y la evaluación de técnicas de autenticación que emplean ZPK.
- Aplicaciones en seguridad en la nube y *blockchain*: se priorizaron trabajos que investigaran la aplicación de ZPK en el contexto de la seguridad en la nube y en tecnologías de *blockchain*; se destacan casos de uso como la autenticación en sistemas distribuidos, la protección de la privacidad en transacciones y la integridad de los datos.

La metodología empleada en esta investigación se dividió en dos enfoques complementarios. En primer lugar, se realizó una revisión bibliográfica, en la que se exploró la literatura existente sobre métodos de autenticación basados en ZKP en el contexto de la seguridad en la nube y *blockchain*. Esta revisión proporcionó una visión general del estado actual de la investigación en este campo y permitió señalar posibles áreas de mejora o innovación.

En segundo lugar, se aplicó un enfoque analítico-sintético que implicó el análisis detallado de los principios teóricos subyacentes a las ZPK y de su aplicación en autenticación criptográfica. A través de una revisión bibliográfica, se profundizó en conceptos clave y se sintetizó

el conocimiento adquirido para identificar posibles áreas de mejora o nuevas direcciones de investigación. Este enfoque combinó la comprensión teórica con la identificación de aplicaciones prácticas, lo que proporcionó una base sólida para avanzar en el campo de la autenticación basada en criptografía de conocimiento cero.

Existen diversos métodos para autenticar a los usuarios que varían en seguridad y complejidad. Estos se pueden agrupar en tres grandes categorías:

1. Autenticación basada en conocimiento

Contraseñas alfanuméricas

Entre los métodos más tradicionales y ampliamente utilizados dentro de la autenticación basada en conocimiento se encuentran las contraseñas alfanuméricas. Son una combinación de caracteres que utilizan el código ASCII para generar una clave secreta de dimensión variable. Son fáciles de implementar y muy comunes en todos los tipos de sistemas. Una contraseña alfanumérica fuerte debe ser aleatoria y larga, por tanto, difícil de recordar. Por la naturaleza de los procesos mnemotécnicos cuando los usuarios tienen que establecer o recordar muchas contraseñas, que utilizan en sistemas distintos, suelen emplear claves parecidas o con frases sencillas de recordar (Rodríguez Valdés et al., 2018). Este método, aunque es el más común, es en extremo vulnerable a varios tipos de ataques, entre ellos:

- *Phishing*: donde los atacantes engañan a los usuarios para que revelen sus contraseñas mediante sitios web falsificados o correos electrónicos maliciosos.
- Ataques de diccionario: que utilizan listas predefinidas de palabras comunes y combinaciones alfanuméricas para intentar adivinar la contraseña.
- Ataques de fuerza bruta: que implican probar todas las combinaciones posibles hasta encontrar la contraseña correcta.

One-Time Password (OTP)

En contraste con las contraseñas estáticas, otro método común dentro del mismo grupo de autenticación basada en conocimiento es el *One-Time Password* (OTP). Dicho sistema asegura el uso de

contraseñas con un sistema OTP, es decir, el usuario posee una contraseña por un tiempo determinado y para un uso específico. Esta solución es usada en general para el proceso de autenticación de inicio para los accesos externos, los cuales se realizan mediante VPN (Red Privada Virtual) (Mendoza Arteaga et al., 2020). Aunque los sistemas OTP ofrecen una capa adicional de seguridad al generar contraseñas temporales, que caducan después de un solo uso, también presentan vulnerabilidades que los atacantes pueden explotar. Por ejemplo, si un atacante logra interceptar el canal de comunicación (como mensajes SMS o correos electrónicos), a través del cual se envía el OTP, podría utilizar el código antes de que expire y acceder a la cuenta o sistema objetivo. Además, los dispositivos móviles utilizados para recibir OTP pueden ser comprometidos por *malware*, lo que permitiría a los atacantes capturar los códigos en tiempo real.

Otra vulnerabilidad significativa es el *phishing* dirigido, donde un atacante engaña al usuario para que revele su OTP en un sitio web falso o por medio de una aplicación maliciosa. De esta forma el atacante puede autenticarse como si fuera el usuario legítimo. Estos riesgos resaltan la necesidad de implementar medidas de seguridad adicionales.

Autenticación Gráfica

Los Sistemas de Autenticación Gráfica son usados tanto para la autenticación de usuarios en un sistema como para la generación de claves para el uso en algoritmos criptográficos (Shendage, S. et al., 2014). Las contraseñas gráficas pueden estar formadas por la combinación de fotos, imágenes o iconografías. Las características de la imagen producen un espacio de claves mucho mayor. La eficiencia de estos sistemas se basan en la gran capacidad de los seres humanos de recordar patrones en imágenes, en vez de memorizar conjuntos de caracteres de gran longitud y complejidad (Rodríguez Valdés et al., 2018).

En general, estos sistemas poseen ventajas demostradas sobre los métodos tradicionales de introducción de texto y conservan espacios de claves considerables. Sin embargo, la autenticación gráfica también presenta algunas desventajas significativas. Uno de sus principales inconvenientes es que su implementación puede ser más compleja y costosa

que los métodos tradicionales, ya que requieren interfaces gráficas más avanzadas y sistemas de almacenamiento y recuperación de imágenes eficientes. Además, dependen de dispositivos con capacidades gráficas modernas, lo que puede limitar su uso en dispositivos más antiguos o con recursos limitados. El empleo de imágenes personales o identificables puede plantear problemas de privacidad y, aunque los espacios de claves son mayores, los ataques de fuerza bruta siguen siendo una amenaza. La interfaz de usuario puede ser menos intuitiva para algunos usuarios, y la compatibilidad entre diferentes sistemas y plataformas puede ser un problema.

2. Autenticación basada en posesión

Tókenes físicos

Además de lo que el usuario conoce, la autenticación basada en posesión se apoya en lo que el usuario posee, como es el caso de los tókenes físicos. Un token es un dispositivo de almacenamiento que contiene una clave secreta que se utiliza en el proceso de autenticación (Mendoza Arteaga et al., 2020). Los tókenes son emitidos y certificados por una entidad de confianza, como una autoridad de certificación, que valida la autenticidad y la integridad de la información contenida en el token, lo que le añade una capa adicional de seguridad.

De manera general, estos tókenes se presentan como tarjetas inteligentes o magnéticas, pero pueden verse en muchas otras como dispositivos digitales o memorias USB. Estas tarjetas contienen un microchip capaz de almacenar información y realizar funciones criptográficas, lo que permite su uso en una amplia gama de aplicaciones, desde el comercio electrónico hasta el control de acceso físico y lógico (Mendoza Arteaga et al., 2020). En muchos países, las tarjetas inteligentes se han convertido en una parte integral de los documentos de identidad, lo que facilita no solo la identificación personal sino también la autenticación segura en servicios en línea y presenciales.

Además de los riesgos de pérdida o robo, los tókenes (como cualquier dispositivo físico) pueden deteriorarse con el tiempo debido al uso continuo. Este deterioro afecta la funcionalidad del token lo que lo vuelve menos fiable, incluso, en algunos casos, inutilizable.

3. Autenticación basada en inherencia

Biometría

Dentro de los métodos de autenticación, la biometría representa un enfoque basado en lo que el usuario es. Este método consiste en la autenticación de un individuo mediante sus características físicas: forma de su rostro, huellas dactilares, estructura y forma de su voz, patrón de iris y otras características únicas de cada ser vivo. El reconocimiento biométrico consiste en almacenar, medir y comparar los rasgos característicos de un individuo (Mendoza Arteaga et al., 2020).

Los esquemas de autenticación basados en información biométrica son alternativas a las contraseñas alfanuméricas. Para su implementación es necesario, en todos los casos, utilizar elementos de *hardware* especializado. Este requisito eleva los costos y la complejidad de los sistemas, lo que puede ser una barrera para su empleo en entornos donde los recursos son limitados. Además, la precisión y confiabilidad de estos dispositivos pueden verse afectadas por factores externos, como la calidad del *hardware*, las condiciones ambientales o el estado físico del usuario. La biometría puede ser conceptualmente difícil de falsificar, pero fácil de suplantar. Esto no significa que puedan ser copiados o duplicados con la misma facilidad que una contraseña, pero la recolección de datos biométricos (como huellas dactilares o imágenes faciales) puede realizarse sin el conocimiento del usuario en situaciones de alta tecnología. El rendimiento de estos sistemas también puede verse comprometido por circunstancias como la salud del usuario, estrés y otros factores que pueden hacer el proceso extenso e incómodo (Rodríguez Valdés et al., 2028).

Autenticación Multifactor (MFA)

Un solo factor de autenticación no es seguro, por lo que en la actualidad la autenticación multifactor (MFA) es la tendencia para verificar que los usuarios sean legítimos (Mendoza Arteaga et al., 2020). Este proceso de seguridad permite que el usuario corrobore su identidad mediante la confirmación de dos factores de autenticación diferentes (clave estática y clave dinámica) (Espinoza Tinoco et al., 2022). Por ejemplo, un sistema puede requerir una contraseña (algo que el usuario sabe) y un código enviado a su teléfono (algo que el usuario tiene).

La MFA aumenta de forma significativa la seguridad al requerir varios niveles de autenticación.

Este tipo de autenticación presenta vulnerabilidades relacionadas con la forma en que se implementa. Si un sitio web utiliza un método de autenticación de dos factores basado en SMS, los atacantes pueden clonar la tarjeta SIM del usuario y redirigir los mensajes de texto a su propio dispositivo. Además, si se utiliza una aplicación de autenticación, el dispositivo del usuario puede verse comprometido por *malware* o virus, lo que permite a los atacantes acceder a los códigos de seguridad generados por la aplicación (Marmolejo-Corona et al., 2023).

La implementación de MFA puede crear un conflicto entre la comodidad y la seguridad. Para muchos usuarios, la necesidad de completar pasos adicionales para autenticarse puede ser percibida como incómoda y engorrosa. Lo anterior puede llevar a una resistencia a su utilización, o incluso a la desactivación de estas medidas si se les da la opción. Además, no todos los servicios en línea son compatibles con MFA, lo que obliga a los usuarios a elegir entre seguridad y conveniencia. Esta incompatibilidad también puede generar dificultades técnicas para configurar o utilizar MFA, especialmente en dispositivos y aplicaciones obsoletos.

Otro desafío significativo de la MFA es la dependencia de proveedores externos para la entrega de códigos de autenticación o para la validación de datos biométricos. Esta subordinación introduce un punto de fallo adicional en el proceso de autenticación. Si el servicio externo es comprometido, la autenticación del usuario también puede estar en riesgo, esto expone al sistema a posibles brechas de seguridad.

Certificados digitales

Entre otros métodos más recientes se encuentra el certificado digital. Este consiste en un documento electrónico que verifica la identidad de una persona u organización en Internet y utiliza tecnología criptográfica para proteger la privacidad y la seguridad de las comunicaciones en línea. Es emitido por una autoridad de certificación responsable

de verificar la identidad del registrante del certificado y garantizar la autenticidad de la clave pública asociada con el certificado. Permite cifrar y descifrar datos, firmar documentos de forma digital y autenticar a usuarios en transacciones y servicios en línea (Marmolejo-Corona et al., 2023).

Aunque son esenciales para la seguridad en línea, los certificados digitales pueden ser vulnerables si se compromete la clave privada asociada. En tal caso, un atacante puede suplantar la identidad del usuario, lo que conlleva graves riesgos de seguridad. Además, tienen una fecha de caducidad y si no se renuevan a tiempo puede interrumpir servicios críticos. La confianza en las autoridades de certificación (CA, por sus siglas en inglés) es esencial, no obstante, si una CA es vulnerada todo el sistema de certificación resulta afectado. Por último, cumplir con normativas y estándares regulatorios implica una gestión constante y puede ser costoso para las organizaciones.

Autenticación mediante el uso de pruebas de ZKP

A pesar de los avances en métodos de autenticación (como la multifactorial y el uso de certificados digitales), persisten desafíos relacionados con la seguridad y la privacidad. Los métodos tradicionales, aunque eficaces en muchos casos, siguen siendo vulnerables a ataques sofisticados y a la exposición de información sensible. Es aquí donde las Pruebas de Conocimiento Cero emergen como una solución innovadora que aborda estos problemas de manera única.

Las ZKP permiten a una parte (*prover*) demostrar a otra (*verifier*) que posee cierto conocimiento o ha realizado una operación sin revelar ninguna información adicional sobre ese conocimiento o el proceso en sí. Este enfoque resuelve una de las limitaciones clave de los métodos de autenticación tradicionales: la necesidad de compartir información sensible para validar la identidad del usuario.

La autenticación mediante ZKP se encuadra, sobre todo, en los métodos basados en conocimiento, ya que el usuario demuestra conocer cierta información sin revelar cuál es esa información. En este contexto, el conocimiento es lo que permite la autenticación, pero a diferencia de los métodos tradicionales (como las contraseñas), el

protocolo de ZKP garantiza que el conocimiento no sea compartido con el sistema de autenticación, lo que añade una capa de seguridad crucial (Fenzi, 2019).

Autenticación en la nube

La computación en la nube ofrece acceso rápido y económico a recursos informáticos como almacenamiento y aplicaciones, lo que la hace atractiva para pequeñas y medianas empresas. Sin embargo, este modelo presenta graves desafíos de seguridad como problemas de privacidad, falta de transparencia, filtraciones de datos, y robos de identidad. La naturaleza conectada y dinámica de la nube aumenta la vulnerabilidad a ataques, ya que las defensas tradicionales no son suficientes para proteger este entorno.

Los servicios en la nube requieren que los usuarios almacenen información sensible, esto expone los datos a posibles accesos no autorizados, sobre todo cuando se utilizan métodos de autenticación débiles. Además, las múltiples copias de datos y servicios en la nube generan un entorno complejo que puede resultar en problemas de gestión y seguridad, y aumenta el riesgo de ataques como el secuestro de sesiones, que consiste en obtener credenciales durante el inicio de sesión (Bermúdez, M. et al., 2020).

Dado este panorama, se requiere un enfoque robusto de seguridad en la nube que trascienda las prácticas tradicionales. Aquí es donde las ZKP se presentan como una solución eficaz. En este contexto, las ZKP permiten verificar la identidad de los usuarios sin revelar información sensible, lo que mitiga los riesgos asociados a la exposición de datos. Su implementación proporciona una capa adicional de seguridad al evitar que los atacantes puedan acceder a información crítica, aun cuando logren interceptar las comunicaciones. Esto no solo protege la integridad y confidencialidad de los datos, sino que también aborda de manera efectiva las principales preocupaciones de seguridad asociadas con la computación en la nube.

Pruebas de conocimiento cero (ZKP)

El protocolo criptográfico ZKP permite compartir y verificar información sin exponer datos sensibles. En esencia, dicho protocolo

diseña un procedimiento en el que un agente (denominado probador) afirma que algo es cierto, mientras que otro agente (el verificador)comprueba que efectivamente lo es, sin necesidad de conocer toda la información. Por tanto, es un sistema para demostrar el conocimiento de algo que ha sucedido sin que la prueba exija revelar todos los datos (Guillou & Quisquater, 1988).

Lo anterior se logra mediante la generación de una prueba por parte del probador que cumple con un conjunto específico de criterios, los cuales el verificador puede utilizar para confirmar la afirmación sin aprender nada más sobre la declaración. En consecuencia, los ZKP permiten minimizar y limitar el acceso a los datos en contextos distribuidos, como los servicios de Internet en general o la computación en la nube (Vilchez Moya et al., 2023).

La criptografía de conocimiento cero se basa en el concepto de «un engañador convincente», donde una entidad, llamada el verificador, puede estar convencida de que otra entidad, llamada el probador, posee cierta información sin aprender nada más sobre ella. Esto se logra mediante la interacción entre el verificador y el probador, en la que se realiza una serie de desafíos y respuestas que permiten verificar la validez de la afirmación sin revelar detalles adicionales (Hernández, D. et al., 2020).

A continuación, se presenta un ejemplo: Alice y Bob se enfrentan a un desafío particular, Bob es daltónico y no puede diferenciar los colores rojo y verde. Sin embargo, Alice tiene en sus manos dos círculos idénticos que difieren solo en el color: uno es rojo y el otro es verde. El objetivo de Alice es demostrarle a Bob que los círculos son diferentes, sin revelar más información. El probador es Alice y el verificador es Bob.

1. Alice muestra los dos círculos a Bob y le explica que son diferentes, pero Bob no puede diferenciarlos.
2. Para comenzar la demostración, Alice le pide a Bob que se ponga ambos círculos a la espalda y le indica que muestre uno de los círculos y luego lo vuelva a esconder.

3. A continuación, Alice le dice a Bob que tiene la opción de volver a mostrar el mismo círculo que mostró inicialmente o cambiarlo por el otro círculo.
4. Cada vez que Bob muestra un nuevo círculo, Alice le indica si cambió el círculo o no.
5. Alice adivina correctamente si Bob cambió o no el círculo en cada turno sucesivo.
6. A medida que Alice adivina de manera correcta, se vuelve cada vez más probable que los círculos sean diferentes.
7. A través de esta interacción y los aciertos consecutivos de Alice, Bob se convence de que los círculos son diferentes sin que Alice tenga que revelar explícitamente por qué.

Una prueba de conocimiento cero debe satisfacer tres propiedades:

- Completitud: si la declaración es válida, entonces el verificador honesto (Bob) estará convencido de este hecho por un probador honesto. En consecuencia, Bob aceptará la identificación de Alice con probabilidad uno.
- Solidez: si la declaración es inválida, ningún probador tramposo puede convencer al verificador honesto de que es verdadera. En consecuencia, el verificador (Bob) rechazará la prueba con una probabilidad de al menos la mitad.
- Conocimiento cero: si la declaración es verdadera, ningún verificador tramposo puede descubrir nada nuevo aparte de esta información (Ezziri & Khadir, 2020).

Existen diferentes protocolos de conocimiento cero que comparten el principio de demostrar que se conoce algo sin tener que decir con exactitud qué es lo que se sabe. En el protocolo de conocimiento cero interactivo, el verificador y el demostrador interactúan entre sí múltiples veces antes de que el verificador pueda estar seguro de que el demostrador conoce la información requerida. El protocolo se fundamenta sobre la base de considerar que, si el usuario realmente conoce la información, entonces debería ser capaz de responder ciertas preguntas o realizar ciertas operaciones con éxito, mientras que un impostor no sería capaz de hacerlo. En consecuencia, las ZPK interactivas tienen aplicaciones en diversas áreas, como la autenticación, la

verificación de identidad, la protección de la privacidad y la seguridad en sistemas distribuidos (Henry, R., 2014).

En cambio, en el Protocolo de Conocimiento Cero No Interactivo (NIZK) el demostrador envía una sola prueba al verificador, sin que sea necesaria ninguna interacción adicional. En este tipo de protocolos (como la prueba de Schnorr), la parte que conoce el secreto interactúa con otra parte (el verificador) para demostrar que conoce el secreto sin revelarlo y sin necesidad de interactuar entre ellos. Algunos casos de sistemas basados en NIZK son las criptomonedas como Zcash y Monero, y las soluciones de autenticación de dos factores basadas en criptografía de clave pública como la autenticación de dos factores universal (U2F).

Otro tipo de protocolo de conocimiento cero aplica la teoría de grafos para demostrar el conocimiento de ciertas relaciones entre los elementos de un grafo. Este protocolo se basa en un desafío-respuesta entre las dos partes. En la primera fase, la parte que quiere demostrar que posee la información (proponente) construye un grafo a partir de dicha información y lo envía a la otra parte (verificador). En la segunda fase, el verificador elige un nodo del grafo y envía una pregunta al proponente pidiéndole que proporcione la información relacionada con ese nodo. En la tercera fase, el proponente responde a la pregunta sin revelar información adicional sobre el grafo. El proceso se repite varias veces sobre diferentes nodos del grafo en cada iteración. Al final, si el proponente puede responder correctamente todas las preguntas del verificador, se considera que ha demostrado que conoce la información sin revelarla.

Por último, cabe mencionar el Protocolo de Conocimiento Cero Basado en Criptografía de Curva Elíptica, un sistema de cifrado muy seguro que utiliza matemáticas complejas para proteger y descifrar información. En este sistema, también se diferencia entre el emisor, el verificador y la prueba. Esta última es el proceso que permite verificar la autenticidad de la información sin revelarla y para realizarla, el emisor y el verificador seleccionan una curva elíptica y un punto de partida en esa curva. El emisor luego realiza una serie de cálculos matemáticos utilizando la información que desea verificar y envía los

resultados de estos cálculos al verificador, quien utiliza estos resultados para determinar si la información es auténtica o no. Este protocolo es muy rápido y eficiente, lo que lo hace ideal para su uso en sistemas en línea y otros entornos de alta velocidad (Wanden-Berghe, C.A., 2013).

Protocolos utilizados en la autenticación

Protocolo de Schnorr

El esquema de autenticación y firma de Schnorr obtiene su seguridad de la dificultad de calcular logaritmos discretos. Para generar un par de claves, primero se eligen dos números primos (p y q), de modo que q sea un factor primo de $p-1$. Luego, se elige una que no sea igual a 1, tal que $a^q \equiv 1 \bmod p$. Todos estos números pueden ser comunes a un grupo de usuarios y pueden ser publicados libremente. Para generar un par de claves específicas (clave pública/clave privada) se elige un número aleatorio menor que q. Esta es la clave privada s. Luego, se calcula $v \equiv a^{-s} \bmod p$. Esta es la clave pública (Schneier, 1996). Protocolo de Autenticación de Schnorr (Véase Figura 1):

1. Alice elige un número aleatorio r menor que q y calcula $x = a^r \bmod p$. Esta es la etapa de preprocesamiento y se puede realizar mucho antes de que Bob esté presente.
2. Alice envía x a Bob.
3. Bob envía a Alice un número aleatorio e entre 0 y $2^t - 1$.
4. Alice calcula $y = (r + s \cdot e) \bmod q$ y envía a Bob.
5. Bob verifica que $x = a^y \cdot v^e \bmod p$.

Esto es porque $x = a^{r+s \cdot e} \cdot (a^{-s})^e = a^r \bmod p$.

La seguridad se basa en el parámetro t. La dificultad para romper el algoritmo es de aproximadamente 2^t. Schnorr recomendó que p tenga alrededor de 512 bits, q alrededor de 140 bits, y t sea 72 bits.

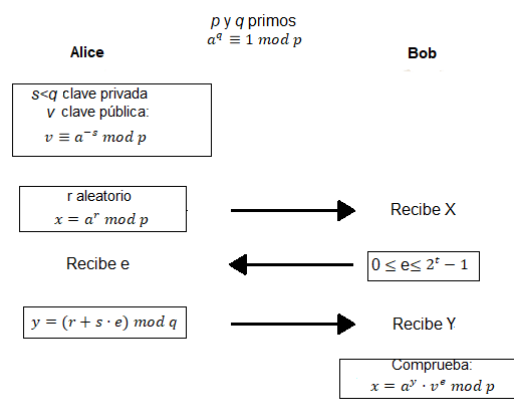

Figura 1. Ronda Protocolo de Schnorr

Dado que el cálculo del logaritmo discreto *s*, a partir de $v \equiv a^{-s} \, mod \, p$ es computacionalmente difícil, el protocolo asegura que un atacante no puede derivar la clave privada a partir de la clave pública.

Protocolo de Identificación Fiege-Fiat-Shamir

El protocolo creado por Uriel Fiege, Amos Fiat y Adi Shamir en 1988 es una de las implementaciones más comunes de una ZPK. El protocolo Fiat-Shamir se basa en la dificultad de la raíz cuadrada. El reclamante demuestra su conocimiento de una raíz cuadrada módulo un gran número *n* (Hasan, J., 2019).

El Esquema de Identificación Feige-Fiat-Shamir incluye dos entidades (el probador y el verificador), de acuerdo con el principio de todas las ZKP. El probador tiene un token secreto y busca autenticación, por lo que debe demostrar a otra entidad (el verificador) que lo autentique en función del token secreto del probador a través de una secuencia de problemas, sin que el verificador conozca el token secreto del probador.

Configuración única

Un centro de confianza publica un módulo $N = p \cdot q$ que es el producto de dos números primos. Cada posible reclamante (proponente) selecciona un secreto *s* que sea coprimo con N y publica $v = s^2 \, mod \, N$ como su clave pública (Figura 2).

1. Público: el módulo N y $v = s^2 \, mod \, N$.
2. Alice selecciona r y envía a Bob $x = r^2 \, mod \, N$.
3. Bob elige e ∈ {0,1}.
4. Alice envía $y = r \cdot s^e \, mod \, N$.
5. Bob debe verificar: $y^2 = x \cdot v^e \, mod \, N$.

Esto es porque $y^2 = r^2 \cdot s^{2e} = r^2 \cdot (s^2)^e = \; x \cdot v^e \, mod \, N$.

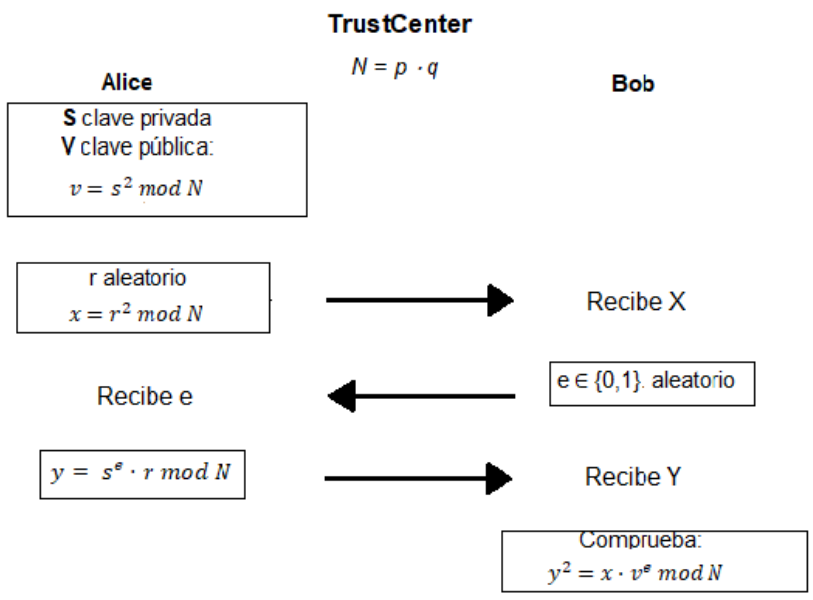

Figura 2. Ronda Fiege-Fiat-Shamir

Por ejemplo, si elegimos $e = 1$, entonces:

1. Público: el módulo N y $v = s^2 \ mod \ N$.
2. Alice selecciona r y envía a Bob $x = r^2 \ mod \ N$.
3. Bob elige $e = 1$.
4. Alice envía $y = r \cdot s \ mod \ N$.
5. Si $y^2 = x \cdot v \ mod \ N$, entonces Bob lo acepta, lo que significa que Alice pasa esta iteración del protocolo.

Por lo tanto, Alice debe conocer s en este caso.

El protocolo se basa en la dificultad de calcular raíces cuadradas modulares sin conocer s. Esto se garantiza por las propiedades derivadas del Teorema de Euler que aseguran la dificultad de invertir las operaciones modulares y la dispersión adecuada de los resultados. Al repetir este proceso varias veces, Bob puede estar seguro de que Alice conoce el secreto s, aunque nunca lo revele de forma directa. Cada iteración refuerza la confianza de Bob en que Alice realmente conoce el secreto.

Protocolo Guillou-Quisquater

Feige-Fiat-Shamir fue el primer protocolo práctico basado en identidad. Por su parte, el protocolo de Guillou-Quisquater se diseñó para disminuir el número de iteraciones y acreditaciones por iteración empleado en el primero al aumentar el trabajo computacional. Para algunas implementaciones, como las tarjetas inteligentes, un gran número de iteraciones es menos que ideal. Los intercambios con el mundo exterior consumen tiempo y el almacenamiento requerido para cada acreditación puede sobrecargar los recursos limitados de la tarjeta.

Louis Guillou y Jean-Jacques Quisquater desarrollaron un algoritmo de identificación de conocimiento cero más adecuado para aplicaciones como estas. Los intercambios entre Alice y Bob, y las acreditaciones paralelas en cada intercambio se mantienen al mínimo absoluto: solo hay un intercambio de una acreditación por cada prueba. Para el mismo nivel de seguridad, la computación requerida por Guillou-Quisquater es mayor que la de Feige-Fiat-Shamir por un factor de tres. Y al igual que Feige-Fiat-Shamir, este algoritmo de

identificación puede convertirse en un algoritmo de firma digital (Schneier, B., 1996).

El protocolo de identificación Guillou-Quisquater es un esquema de ZPK diseñado para optimizar la autenticación en dispositivos de seguridad (como microprocesadores o tarjetas inteligentes), lo que minimiza tanto la transmisión como los requisitos de memoria, a cambio de un aumento en el trabajo computacional.

Una vez más, sea n el producto de dos números primos p y q. Sea v coprimo con $\varphi(n) = (p - 1)(q - 1)$.

Sea J la clave pública de Alice, entonces, se define su clave privada B tal que $J \cdot B^v \equiv 1 \ mod \ n$.

Alice envía a Bob sus credenciales, J. Ahora, quiere probarle a Bob que esas credenciales son suyas. Para hacerlo, debe convencer a Bob de que conoce B. En la Figura 3 se presenta el protocolo:

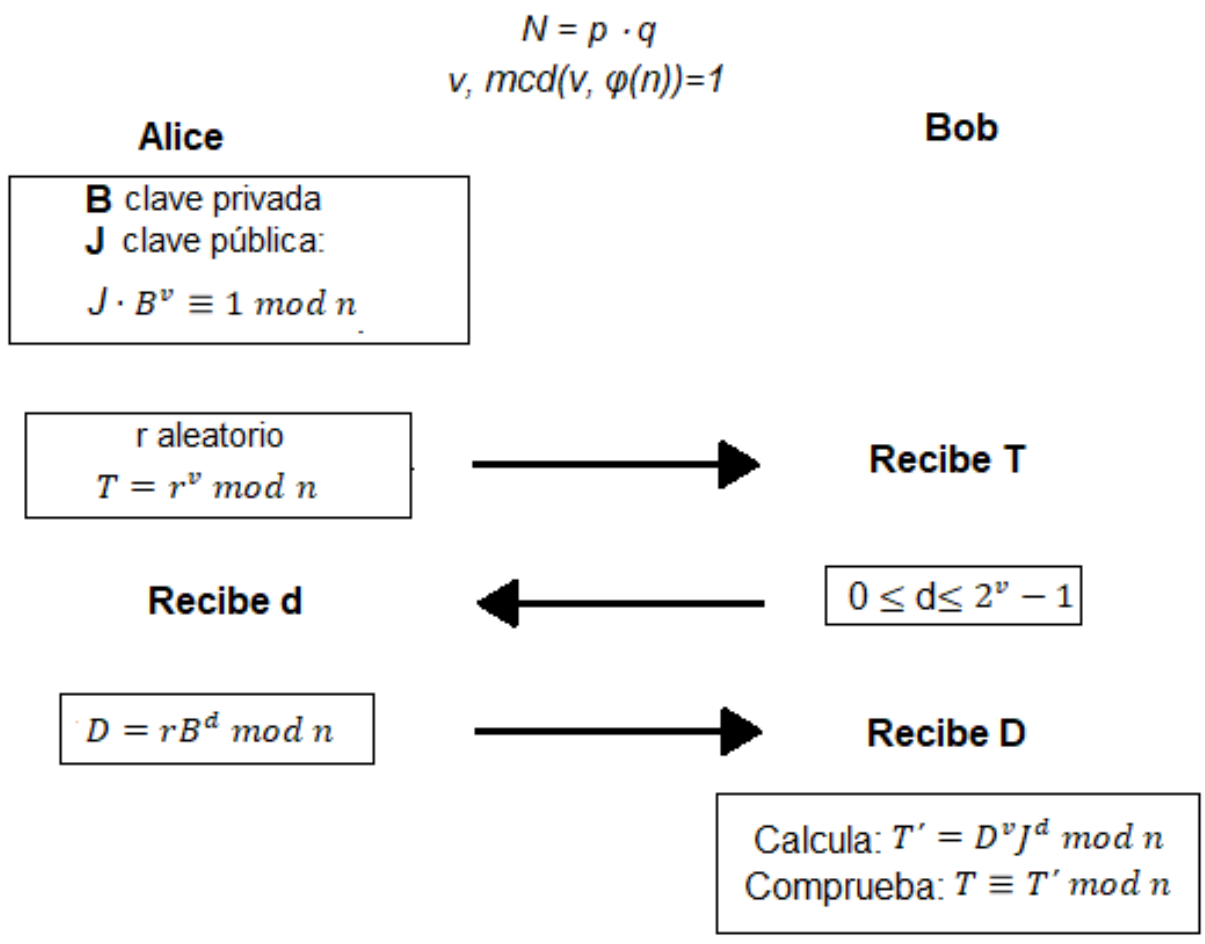

Figura 3. Ronda Guillou-Quisquater

1. Alice elige un número entero aleatorio r, tal que r esté entre 1 y $n-1$. Calcula $T = r^v \ mod \ n$ y lo envía a Bob.

2. Bob elige un número entero aleatorio, e, tal que e esté entre cero y $v-1$. Envía e a Alice.

3. Alice calcula $D = rB^d \ mod \ n$ y lo envía a Bob.

4. Bob calcula $T' = D^v J^d \ mod \ n$. Si $T \equiv T' \ mod \ n$, entonces la autenticación tiene éxito.

Las matemáticas no son tan complejas:

$$T' = D^v J^d = (rB^d)^v J^d = r^v D^{dv} J^d = r^v (J \cdot B^v)^d = r^v \equiv T \bmod n.$$

Dado que B fue construida para satisfacer $J \cdot B^v \equiv 1 \bmod n$ (Schneier, 1996) y (Guillou & Quisquater, 1988).

La seguridad de este protocolo reside en el parámetro *v*, ya que define el rango de valores entre los cuales Bob puede elegir su desafío *d* y, en consecuencia, la probabilidad de que un impostor pueda adivinarlo. Si Eve lograra adivinar el valor de *d*, podría hacerse pasar por Alice incluso sin conocer su clave privada *s*. Podría elegir un valor y aleatorio de antemano que sería el que envíe en el tercer paso y definir $x = y^v J^d$. En el primer paso en consecuencia; de esta manera, la verificación de Bob en el paso final sería trivialmente cierta. Recordar que, aunque Eve no conoce la clave privada de Alice, tiene acceso a su clave pública (Vilchez Moya et al., 2023).

Seguridad en estos protocolos

Uno de los ataques más comunes en los métodos de autenticación es el de *replay*, que consiste en reutilizar una comunicación exitosa anterior entre dos partes para suplantar la identidad de una de ellas mediante el reenvío de mensajes intercambiados. Aunque muchos protocolos de autenticación, como Kerberos, son vulnerables a este ataque, los protocolos de ZKP son resistentes debido a que cada ejecución es única gracias a la aleatorización utilizada. En el caso de los ZKP interactivos, también son negables a terceros que no participaron en la prueba, una propiedad que no se mantiene en los ZKP no interactivos.

Gracias a su naturaleza de conocimiento cero, la seguridad en los protocolos de autenticación basados en ZKP no se deteriora con el tiempo y es resistente a la interceptación de comunicaciones (*eavesdropping*) y al ataque de *chosen-text*. Este último ataque, común en esquemas de autenticación de tipo *challenge-response*, busca extraer información del secreto del probador seleccionando desafíos específicos. Sin embargo, los ZKP evitan que un verificador deshonesto obtenga información porque las respuestas del probador siempre contienen aleatoriedad. Además, los ZKP previenen

efectos no deseados, como la firma digital involuntaria de un archivo, que pueden ocurrir cuando se reutiliza la misma clave RSA para autenticación y firma digital.

El ataque *man-in-the-middle*, donde un atacante intercepta y modifica mensajes entre dos partes, es ineficaz contra los protocolos ZKP ya que el secreto nunca se transmite por el canal de comunicación. Además, ZKP protege contra la suplantación de identidad (un riesgo en esquemas de autenticación débiles basados en contraseñas), pues nunca revela el secreto ni información relacionada al verificador. En cuanto al *phishing*, ZKP ofrece una defensa sólida, y evita que los usuarios revelen sus credenciales a través de técnicas engañosas como correos fraudulentos o sitios web clonados. Dos protocolos ZKP propuestos proporcionan autenticación sin necesidad de compartir contraseñas, superando otros esquemas actuales.

En resumen, los protocolos de autenticación ZKP se destacan como una de las opciones más seguras frente a esquemas tradicionales como la autenticación débil con contraseñas y la autenticación fuerte de *challenge-response*, gracias a su resistencia a numerosos ataques conocidos. Las debilidades de los protocolos ZKP suelen surgir de deficiencias en su implementación, dado que sus fundamentos teóricos son sólidos y robustos (Bukovits, 2023).

Conclusiones

El presente estudio destacó la relevancia y el impacto de las Pruebas de Conocimiento Cero (ZKP) en el ámbito de la autenticación criptográfica, en especial en contextos como la seguridad en la nube. La revisión exhaustiva de los principales protocolos y métodos de ZKP ha revelado que estas técnicas juegan un papel crucial en la mejora de la privacidad y la seguridad en sistemas distribuidos, lo que ofrece garantías robustas sin comprometer la confidencialidad de la información.

La implementación de ZKP representa un avance significativo en la autenticación segura y permite verificar la validez de una transacción o identidad sin necesidad de revelar información sensible. La flexibilidad y adaptabilidad de estos métodos son valiosas en un

entorno donde la protección de datos y la privacidad son de suma importancia. Las diversas aplicaciones de ZKP demuestran el potencial de estas técnicas para fortalecer la seguridad en diversos contextos.

Por último, la protección de la privacidad y el desarrollo de marcos éticos para el uso de tecnologías de autenticación basadas en ZKP son aspectos fundamentales que deben ser considerados. A medida que estas técnicas se integran en aplicaciones más amplias, es esencial explorar enfoques para garantizar la privacidad de los datos y establecer directrices éticas para su implementación en entornos sensibles. Estas áreas representan importantes direcciones para futuras investigaciones y desarrollo en el campo de la autenticación criptográfica.

Referencias bibliográficas

Bermúdez León, M. J. (2020). *Métodos de seguridad, computación en la nube*. Repositorio USAM. http://repositorio.usam.ac.cr/xmlui/handle/11506/2202

Bukovits, N. A. (2023). *Zero Knowledge proofs y sus aplicaciones prácticas* [Tesis de grado, Universidad de Buenos Aires]. https://www.sidalc.net/search/Record/tpos:1502-2432_BukovitsNA/Description

Espinoza Tinoco, L., Barriga Rivera, B., Izurieta Navarrete, J., & Morales Alarcón, C. H. (2022). Seguridad informática aplicando la Autentificación por Doble factor para la plataforma HomeOfi. *Ciencias Técnicas y Aplicadas*, *8*(3). https://dialnet.unirioja.es/servlet/articulo?codigo=8637935

Ezziri, S., & Khadir, O. (2020). A Zero-Knowledge Identification Scheme Based on the Discrete Logarithm Problem and Elliptic Curves. En F. Saeed, T. Al-Hadhrami, E. Mohammed, & A. Al-Sarem (Eds.), *Advances on Smart and Soft Computing* (Vol. 1188, pp. 13-23). Springer. https://doi.org/10.1007/978-981-15-6048-4_3

Fenzi, G. (2019). *Zero Knowledge Proofs Theory and Applications* [Tesis de maestría, University of St. Andrews]. https://info.cs.st-andrews.ac.uk/student-handbook/files/project-library/cs4796/2020/gf45-Final_Report.pdf

Guillou, L. C., & Quisquater, J.-J. (1988). A practical zero-knowledge protocol fitted to security microprocessor minimizing both transmission and memory. En C. G. Günther (Ed.), *Advances in Cryptology — EUROCRYPT '88* (Vol. 330, pp. 123-128). Springer. https://link.springer.com/chapter/10.1007/3-540-45961-8_11

Hasan, J. (2019). Overview and Applications of Zero Knowledge Proof (ZKP). *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *8*(5). https://www.academia.edu/download/63348598/Overview-and-Applications-of-Zero-Knowledge-Proof-ZKP20200518-20172-1ev7qkn.pdf

Henry, R. (2014). *Efficient Zero-Knowledge Proofs and Applications* [Tesis doctoral, University of Waterloo]. UWspace. https://uwspace.uwaterloo.ca/handle/10012/8621

Hernández Fernández, D. (2020). *Pruebas de conocimiento cero* [Tesis de grado, Universidad de Salamanca]. GREDOS. https://gredos.usal.es/handle/10366/152749

Marmolejo-Corona, I. V., Bautista-Aguilar, F. A., Santiago-González, Y. F., & Serrano-Manzano, G. A. (2023). Seguridad en Sistemas de Autenticación: Análisis de Vulnerabilidades y Estrategias de Mitigación. *Xikua Boletín Científico de la Escuela Superior de Ciudad Sahagún*, 11(22). https://repository.uaeh.edu.mx/revistas/index.php/xikua/article/view/10802

Mendoza Arteaga, A., Bolaños Burgos, F., Cedeño Sarmiento, C., & Saltos Rivas, W. R. (2020). La importancia de la autenticación multifactor para el usuario final en un entorno financiero. Informática y Sistemas: *Revista de Tecnologías de la Informática y las Comunicaciones*, *4*(1). https://revistas.utm.edu.ec/index.php/Informaticaysistemas/article/download/2347/2560

Quisquater, J.-J., Guillou, L. C., Berson, T. A., Blåken, G., Guillot, P., Loret, E., Loyer, B., Lussis, P., M'Raihi, D., Noro, M., Poupard, C., Pralile, O., Riguidel, M., Tardy-Corfdir, A., & Vallee, B. (1989). How to Explain Zero-Knowledge Protocols to Your Children. En G. Brassard (Ed.), *Advances in Cryptology — CRYPTO' 89 Proceedings* (Vol. 435, pp. 628-631). Springer. https://link.springer.com/chapter/10.1007/0-387-34805-0_60

Rodríguez Valdés, O., Legón, C. M., & Socorro Llanes, R. (2018). Seguridad y usabilidad de los esquemas y técnicas de autenticación gráfica. *Revista Cubana de Ciencias Informáticas, 12*(2). http://scielo.sld.cu/scielo.php?pid=S2227-18992018000500002&script=sci_arttext&tlng=pt

Schneier, B. (1996). *Applied Cryptography, Second Edition: Protocols, Algorthms, and Source Code in C*. John Wiley & Sons. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119183471

Shendage, S. S., Dhainje, P., & Yevale, R. S. (2014). Cued Click Points: Graphical Password Authentication Technique for Security. *International Journal of Computer Science and Information Technologies,* 5(2). https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ec7ee0b0434b9ac467afe5c2a1ce0df35a802b63

Vilchez Moya, C., Bermejo Higuera, J. R., Bermejo Higuera, J., & Sicilia Montalvo, J. A. (2023). Implementation and Security Test of Zero-Knowledge Protocols on SSI Blockchain. *Applied Sciences, 13*(9), 5717. https://doi.org/10.3390/app13095717

Wanden-Berghe Fajardo, C. A. (2023). *Blockchain e inteligencia artificial en el sistema de información contable: la disrupción de la partida triple* [Tesis doctoral, Universidad de Alicante]. RUA. https://rua.ua.es/dspace/handle/10045/135180

INVESTIGACIÓN

Integración de prácticas de seguridad para pipelines DevSecOps en entornos contenerizados

Integration of security practices for DevSecOps pipelines in containerized environments

Ing. Mary Nelsa Bonne Cuza[*], Téc. Meliza León Bencomo[2]

Recibido: 11/2025 | Aceptado: 11/2025 | Publicado: 12/2025

Resumen

La seguridad en el desarrollo de *software* es una preocupación primordial en la era digital actual, especialmente en entornos de integración y entrega continua. A medida que las amenazas cibernéticas evolucionan, es necesario que los equipos de desarrollo integren la seguridad desde las fases iniciales del ciclo de vida del *software*. La investigación se centra en implementar de prácticas DevSecOps, que buscan incorporar la seguridad desde el inicio del *pipeline*. Se emplearon métodos cualitativos y cuantitativos como análisis de casos y revisión documental, para identificar mejores prácticas y desafíos en el campo. Los hallazgos indican que la falta de alineación entre los equipos de desarrollo, operaciones y seguridad es un obstáculo crítico. Sin embargo, la implementación efectiva de herramientas automatizadas mejora de forma significativa seguridad. Las conclusiones destacan la necesidad de fomentar una cultura colaborativa y un

[*] Facultad de Ciberseguridad. Universidad de las Ciencia Informáticas. Carretera a San Antonio Km 2 ½ Torrens. Boyeros. La Habana. Cuba. mary@uci.cu

[2] Dirección de Seguridad Informática. Universidad de las Ciencia Informáticas. Carretera a San Antonio Km 2 ½ Torrens. Boyeros. La Habana. Cuba. mary@uci.cu

enfoque proactivo hacia la seguridad. Se resalta la urgencia de abordar las preocupaciones de seguridad en el desarrollo de *software* y se propone un proceso claro hacia una integración exitosa de prácticas DevSecOps.

Palabras clave: contenerización, entrega continua, integración continua, seguridad

Abstract

Security in software development is a paramount concern in today's digital age, especially in continuous integration and continuous delivery environments. As cyber threats evolve, it is necessary for development teams to integrate security from the early stages of the software lifecycle. The research focuses on the implementation of DevSecOps practices, which seek to incorporate security from the beginning of the pipeline. Qualitative and quantitative methods, such as case analysis and document review, were used to identify best practices and challenges in the field. The findings indicate that the lack of alignment between development, operations and security teams is a critical obstacle. However, effective implementation of automated tools significantly improves the security posture. The findings highlight the need to foster a collaborative culture and a proactive approach to security. It highlights the urgency of addressing security concerns in software development, proposing a clear process towards successful integration of DevSecOps practices.

Keywords: *containerization, continuous delivery, continuous integration, security*

Introducción

La evolución de las prácticas de desarrollo de *software* ha llevado a la creación de metodologías que buscan mejorar la eficiencia y la colaboración entre los equipos involucrados. Una de las innovaciones más significativas en el ámbito es el concepto de DevSecOps, que ha transformado la manera en que los desarrolladores y los equipos de operaciones interactúan y trabajan juntos. Su propósito es optimizar el proceso de entrega de *software*. Esta metodología promueve la

colaboración entre operadores y desarrolladores, con el objetivo de igualar su agilidad hacia un objetivo común (Muñoz, 2023).

La seguridad de la información y el cumplimiento normativo son fundamentales en el desarrollo de *software*. Estadísticas recientes revelan un aumento notable en los costos asociados a las violaciones de datos, así como en la frecuencia y sofisticación de los ataques dirigidos a las aplicaciones y sistemas. La gestión de vulnerabilidades se ha convertido en un desafío urgente, con tiempos de reparación que superan los 200 días, según datos del Instituto Nacional de Estándares y Tecnología (NIST, por sus siglas en inglés) (Madnick, 2023).

El panorama muestra una tendencia creciente hacia la automatización y la mejora continua en las operaciones de *software*. Entre 2015 y 2025 se ha observado una adopción significativa de DevSecOps, cuyo crecimiento exponencial refleja la madurez progresiva de las prácticas de seguridad integradas. En este contexto, DevSecOps se consolida como el enfoque clave para abordar los desafíos de seguridad. La metodología se establece como una necesidad imperativa para mitigar riesgos y asegurar la integridad de las aplicaciones (Chandramouli et al., 2024; *GitLab*, 2025). Para mejorar la velocidad de implementación, la aplicación de parches, el escalado y la eficiencia en el desarrollo del *software*, se utilizan las tecnologías tales como Docker y Kubernetes. En esta investigación las autoras se centraron en la tecnología de contenerización Kubernetes.

Kubernetes permite organizar y gestionar el trabajo en DevSecOps de manera efectiva al proporcionar una infraestructura escalable y automatizada que facilita el despliegue de aplicaciones. Muchas organizaciones han adoptado esta tecnología para implementar políticas de seguridad específicas y reducir la superficie de ataque, mejorando la resistencia ante posibles amenazas. La integración de DevSecOps en Kubernetes se materializa a través de un *pipeline* de seguridad continua que abarca múltiples capas de protección (Kubernetes, 2024).

No obstante, el entorno presenta una serie de desafíos críticos relacionados con la seguridad del *pipeline*. Una de las problemáticas más significativas se manifiesta cuando un despliegue reciente se ve comprometido por una vulnerabilidad no detectada en una de

las imágenes del contenedor. La ausencia de estándares específicos y mejores prácticas establecidas impide que el equipo de desarrollo cuente con un marco referencial que dirija la implementación de medidas de seguridad a lo largo del SDLC (*System Development Life Cycle*).

Este vacío normativo retrasa la detección de los problemas y compromete la integridad de la aplicación, lo que afecta de manera negativa la confianza del cliente y la reputación corporativa. La situación genera incertidumbre con respecto a la conformidad regulatoria que podría conducir a sanciones económicas significativas y la pérdida de licencias operativas para el funcionamiento del negocio. La carencia de visibilidad respecto al estado y comportamiento del *pipeline* también representa un desafío crítico. Sin herramientas apropiadas para la supervisión resulta casi imposible detectar anomalías o comportamientos sospechosos en tiempo real. Dicha situación conduce a un aumento en los tiempos de respuesta ante incidentes, lo que afecta gravemente la continuidad operativa. En ausencia de prácticas estandarizadas para gestionar configuraciones seguras, el equipo de desarrollo identifica que muchas aplicaciones son implementadas con configuraciones inseguras por defecto, lo que genera un entorno propenso a ataques cibernéticos y vulnerabilidades susceptibles de ser explotadas por actores maliciosos.

Por lo antes expuesto, las autoras decidieron realizar una investigación sobre las prácticas más idóneas en materia de seguridad que se pueden integrar en el *pipeline* DevSecOps en entornos contenerizados con el fin de garantizar la protección integral del *software* desde su concepción hasta su despliegue en producción.

Materiales y métodos

Para la realización de la propuesta de solución, las autoras tuvieron en cuenta varios métodos científicos que permitieron una mayor organización en el trabajo. Se utilizó el método **análisis-síntesis** para descomponer el problema de investigación en sus componentes principales y, posteriormente, integra estos elementos en la propuesta del procedimiento. Este proceso facilitó la identificación de las in-

terrelaciones entre las herramientas, estándares y mejores prácticas de seguridad utilizadas. El método de **observación** contribuyó a la definición del objeto de estudio y las herramientas empleadas; permitió captar información relevante sobre el tema de investigación. El método de **inducción-deducción** se utilizó para tener una visión clara del procedimiento que se quiere elaborar.

Se tomaron en cuenta los fundamentos necesarios sobre los elementos que intervienen en los *pipelines* DevSecOps en Kubernetes como la entrega e integración continua (CI/CD) en la implementación del *pipeline* DevSecOps, la cual garantiza que cada etapa del ciclo de vida del *software* esté respaldada por medidas robustas (Muñoz, 2023; Xygeni, 2024). Según Kev Zettler (en su publicación de Atlassian, 2024), la implementación de DevSecOps cuenta con cinco etapas cíclicas para su desarrollo: Planificación, Compilación, Prueba, Despliegue y Monitoreo (Atlassian, 2024). Estas etapas se convierten en un marco fundamental para garantizar que las aplicaciones sean seguras y alineadas con los objetivos estratégicos de la organización. Por su parte, Luca Antoniotti, desarrollador de *software*, define, en su tesis de maestría para el Politécnico de Torino, cuatro etapas fundamentales para el *pipeline* DevSecOps con diferencias sutiles respecto a la metodología general: Construcción, Prueba, Entrega y Despliegue (Antoniotti, 2021).

Se consideraron estándares y regulaciones para DevSecOps. Entre las principales se encuentran:

• *ISO/IEC 27001:2022:* proporciona un marco adaptable para alinear la seguridad con los objetivos empresariales (Zhu & Zou, 2021).

• *NIST SP 800-53:* brinda lineamientos específicos para proteger sistemas y datos en entornos dinámicos, como *pipelines* de CI/CD (Zayni & Nassereddine, 2020).

Se efectuó un análisis de soluciones para la integración de estándares y mejores prácticas de seguridad para el *pipeline* DevSecOps en Kubernetes a partir del diseño de un protocolo de búsqueda que facilitó la observación de artículos y estudios relacionados con el tema de investigación y se sentaron las bases para confección de la propuesta de solución.

La solución presentada integra herramientas y tecnologías para la conformación de estándares y mejores prácticas de seguridad para el *pipeline* DevSecOps en entornos contenerizados. Se utilizó Tekton, un marco de automatización de código abierto (específicamente en entornos Kubernetes), en particular para crear sistemas de CI/CD, que permite a los desarrolladores construir, probar e implementar en proveedores de nube y sistemas locales. La elección de Tekton se justifica por su capacidad para proporcionar una plataforma extensible y modular que facilita la automatización de flujos de trabajo (Tekton, 2023).

Resultados y discusión

En el ámbito de la gestión de procesos y controles de seguridad, la estructuración de procedimientos definidos es fundamental para garantizar la consistencia y el cumplimiento de los objetivos establecidos. El procedimiento propuesto en esta investigación se compone de tres fases principales, las cuales agrupan un total de seis pasos que permiten definir, implementar y validar controles de seguridad de manera sistemática, lo que asegura su alineación con estándares reconocidos y la correcta asignación de responsabilidades. En la Figura 1 se muestra el proceso organizado en tres fases principales:

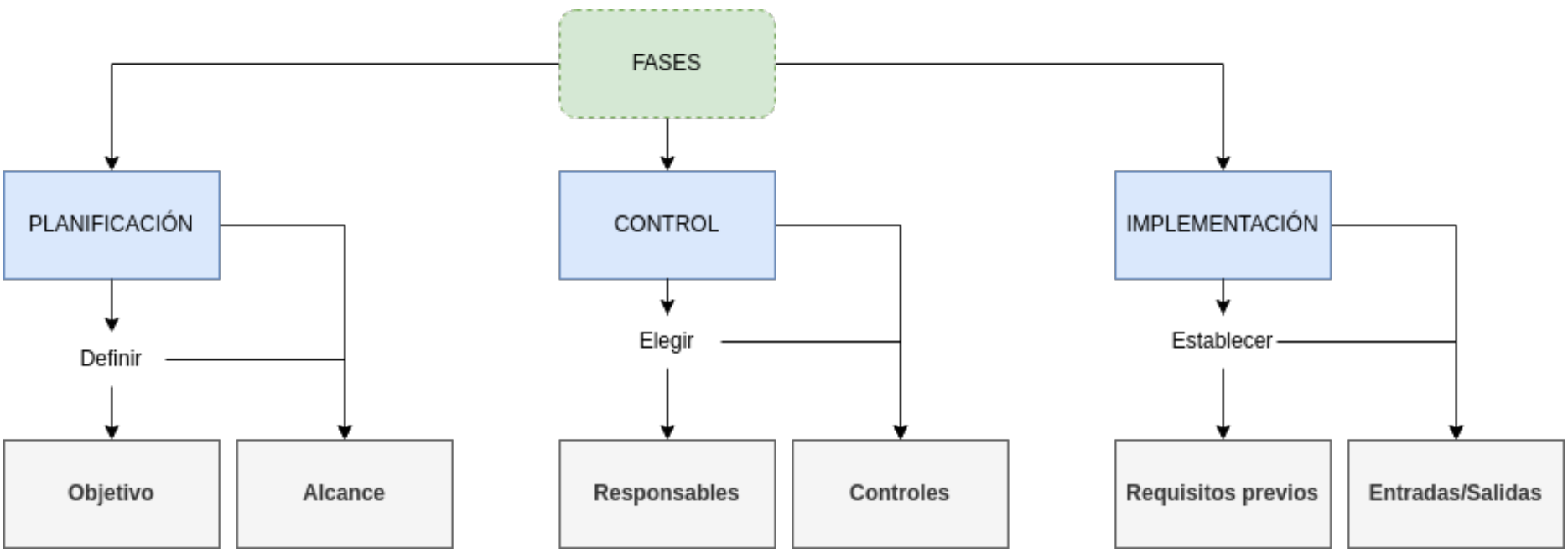

Figura 1. Estructura del procedimiento a desarrollar

1. Fase de Planificación

- Comienza con la definición del objetivo, que establece el propósito principal.
- El alcance se deriva del objetivo y define los límites del procedimiento.

2. Fase de Control

- La selección de controles de seguridad proporciona las medidas necesarias.
- La asignación de responsables garantiza la ejecución adecuada.

3. Fase de Implementación

- Se establecen los requisitos específicos para cada control.
- Por último, se definen las entradas y salidas adaptadas al proceso.

A través de este enfoque, se busca no solo establecer un marco de acción definido, sino también facilitar la adaptabilidad a diferentes contextos organizacionales, integrando requisitos previos, acciones esperadas y mecanismos de retroalimentación. La aplicación metodológica de estos pasos contribuye a la mejora continua, minimiza riesgos y optimiza la eficacia de los controles implementados.

Descripción de las fases para la implementación del procedimiento

1. La **fase de Planificación** es fundamental en cualquier procedimiento, su objetivo principal es establecer de manera clara y precisa qué se desea lograr. Este proceso proporciona una dirección específica y define un propósito concreto que guiará todas las acciones derivadas (véase Cuadro 1).

Cuadro 1. Fase de Planificación del procedimiento

Objetivo del procedimiento	Alcance del procedimiento
• Integrar controles de seguridad automatizados en el *pipeline* de CI/CD mediante estándares reconocidos (*NIST SP 800-53* e *ISO/IEC 27001/2022*), lo que combina la implementación técnica. • El enfoque unificado busca garantizar que la seguridad no sea una fase posterior, sino un proceso automatizado desde la construcción hasta el despliegue, reduciendo riesgos sin comprometer agilidad.	• Alcance multidimensional: La solución propuesta tiene un impacto en tres ámbitos fundamentales: en el empresarial, que abarca todos los niveles organizacionales desde el desarrollo hasta las operaciones; en el técnico, que cubre la infraestructura de Kubernetes y sus componentes; y en el de seguridad, que implementa controles asegurando el cumplimiento de estándares internacionales como la *ISO/IEC 27001:2022* y el *NIST SP 800-53*.

2. El propósito principal de la **fase de Control** es definir un conjunto de medidas de seguridad y verificación que aseguren el cumplimiento riguroso de los estándares establecidos. Implica la selección de mecanismos que monitoreen la calidad del trabajo y la designación de personas específicas y responsables para las tareas asignadas. De esta forma, se garantiza que cada miembro del equipo tenga definido su rol para mantener la integridad del proceso.

Estándares y controles. Criterio de selección

La elección de *ISO/IEC 27001:2022* junto con *NIST SP 800-53* como marcos de seguridad para un *pipeline* DevSecOps se justifica debido a la cobertura integral y enfoque complementario que poseen respectivamente. La base normativa para el procedimiento de integración de seguridad en *pipelines* DevSecOps incluye los siguientes elementos específicos, cada uno con una función generalizada, según las normativas (Wilbur L. Ross, Jr., Secretary, 2020; *ISO/IEC 27001, 2022*)

ISO/IEC 27001/2022

- A.8.1: se enfoca en la gestión y seguridad de dispositivos de usuario finales.
- A.8.2: establece los requisitos para gestionar y controlar los accesos privilegiados dentro de la organización.
- A.12.1.3: establece los componentes necesarios para mantener la capacidad adecuada para garantizar que los servicios sean seguros y estén disponibles.
- A.8.23: implementa medidas para controlar y monitorear el acceso a contenido web.
- A.8.28: se centra en garantizar que el desarrollo de *software* se realice de manera segura.

NIST SP 800-53

- AC-2.7: establece un marco de autenticación único y consistente, lo que permite la integración automática en el *pipeline* mediante la automatización de pruebas de autenticación y la validación continua de identidades. Establece el control de acceso a los recursos.

- AC-3: aporta una capa adicional de seguridad mediante la gestión de acceso, lo que es esencial en entornos DevSecOps donde los cambios son frecuentes y necesitan ser consistentes.
- AC-6: proporciona un enfoque de gestión de privilegios mínimos, lo que permite la automatización de roles y permisos.
- AC-4: establece políticas de red seguras que pueden ser versionadas y aplicadas automáticamente, esto posibilita la segmentación de red y el control de tráfico en cada etapa del *pipeline*.
- AC-10: aporta un enfoque integral de gestión de eventos de seguridad que puede ser automatizado y monitoreado en tiempo real.

Estos controles son particularmente valiosos porque pueden ser automatizados y versionados, lo que permite su integración fluida en el *pipeline* CI/CD, a diferencia de otros controles que requieren intervención manual o que no son de manera directa aplicables a un entorno automatizado. Su implementación conjunta crea una capa de seguridad integral que abarca los aspectos principales de un entorno DevSecOps: autenticación, control de acceso, seguridad de red y monitoreo de eventos.

Estrategia de selección para los controles de seguridad

Para el desarrollo de la investigación se diseñó un protocolo de selección que facilitó el análisis de estándares y controles relacionados con la integración de medidas de seguridad en el *pipeline* DevSecOps en el clúster de k8s que utilizó Tekton. La combinación de estos criterios permitió seleccionar un conjunto de controles que no solo cumplen con los requisitos de seguridad, sino que también son implementables y mantenibles en un entorno DevSecOps moderno, donde la automatización y la velocidad de entrega son fundamentales.

Objetivo del análisis: identificar estándares y controles que describen métodos aplicables para la integración de mejores prácticas para el *pipeline* DevSecOps en Kubernetes. Además, identificar las técnicas de implementación seguras que se pueden integrar al *pipeline* DevSecOps.

Fuente de información: normativas gubernamentales o verificadas en el ámbito nacional. *ISO/IEC 27001* y *NIST SP 800-53*.

Estrategia de búsqueda: Controles que abarquen la implementación de prácticas de seguridad que complementen la robustez de Kubernetes.

Criterios de inclusión y exclusión: Inclusión: capacidad de automatización a través de la posibilidad de implementación mediante *scripts*. Alineación con medidas de seguridad establecidas en el marco regulatorio cubano. Integración con herramientas de automatización utilizadas. Capacidad de evolución con la organización según su adaptabilidad a diferentes entornos empresariales.

Procedimientos de revisión: Selección inicial, Evaluación detallada y Síntesis de resultados.

Responsables de la ejecución del procedimiento: El enfoque DevSecOps trasciende los equipos de desarrollo, operaciones y seguridad. Involucra grupos multidisciplinarios que incluyen equipos de control de calidad (QA), arquitectos de soluciones y especialistas en cumplimiento normativo. La colaboración entre estos grupos resulta fundamental para garantizar la integración de la seguridad en cada fase del ciclo de vida del desarrollo del *pipeline*.

El equipo DevSecOps es el responsable del proceso. Su función principal es garantizar la integridad y protección de los entornos de desarrollo y despliegue. Sus responsabilidades comprenden la definición y mantenimiento de políticas de seguridad, basadas en las normas *NIST SP 800-53*. Es el encargado de la implementación de controles automatizados en la plataforma Tekton, en conformidad con los estándares seleccionados.

Clientes finales: el proceso cuenta con diversos clientes que se dividen en dos grupos: internos, que participan activamente en el proceso DevSecOps; y los externos, que son los beneficiarios finales con una relación independiente del proceso. La implementación de *pipelines* en DevSecOps es fundamental porque integra de forma automática la seguridad en cada etapa del desarrollo de *software*, lo que ayuda a identificar y corregir fallos desde el inicio del proceso. Al automatizar las pruebas de seguridad y la evaluación de vulnerabilidades, las organizaciones pueden entregar *software* más seguro y de mayor calidad de manera más rápida.

Requisitos previos para la implementación de cada control: organizar y aislar los recursos de manera segura permite la creación de una base sólida para el desarrollo y despliegue de los controles seleccionados. A continuación se presentan los requisitos que aseguran que la implementación sea exitosa para crear y gestionar *pipelines*:

Requisito 1: Instalar y configurar Tekton: se debe comenzar con la ejecución de las instrucciones correspondientes a su documentación inicial. El objetivo es asegurar que el código y las dependencias cumplan con estándares de seguridad antes de generar artefactos.

Requisito 2: Configurar el entorno de trabajo

Salidas esperadas: Un *pipeline* de DevSecOps operativo que integre controles de seguridad automatizados, estableciendo un ciclo de retroalimentación continua para la mejora iterativa de las medidas de seguridad a lo largo de todas las etapas, desde el desarrollo hasta la producción.

Implementación en la etapa de construcción del *pipeline*

La etapa de construcción es fundamental para verificar la integridad del código y detectar vulnerabilidades de seguridad. Se lleva a cabo su desarrollo a través de los controles AC-3 y AC-2.7 de *NIST 800-53*.

Control AC-2(7) del *NIST SP 800-53* Esquema Basado en Roles (RBAC)

Objetivo: restringir accesos al *pipeline* de construcción (ejemplo: solo el equipo de DevSecOps puede modificar *Tasks*).

Implementación

• Crear 3 Roles en formato YAML para definir los roles esenciales (un rol de administrador total, un rol de desarrollador y un rol de solo lectura).

Control AC-3 del NIST SP 800-53 (Políticas de red)

Objetivo: garantizar que solo los recursos autorizados puedan interactuar con el *pipeline*, aplicando el principio de mínimos privilegios y seguridad por defecto. Crear *NetworkPolicies* para aislar los *pods*.

Implementación

• Crear *NetworkPolicies* para aislar el *pipeline*.

Implementación en la etapa de prueba del *pipeline*

La implementación del I*SO/IEC 27001:2022* como control de seguridad en el *pipeline* DevSecOps se realizó mediante una arquitectura integrada que comenzó con la definición de dos controles fundamentales: el A.8.1 para la identificación de activos mediante escaneo de secretos, conforme a la legislación cubana que recomienda a las entidades a implementar medidas de seguridad para la información crítica según su Artículo 17 del *Decreto-Ley 370/2018* y el A.8.2 para el control de acceso elevado a sistemas y datos en dependencias, siguiendo las disposiciones del *Decreto-Ley 360/2019* que requiere la gestión de riesgos en infraestructuras críticas según su Artículo 8 (Saber, 2018).

Control A.8.1 y A.8.2 del ISO/IEC 27001-2022 (Gestión de acceso y privilegios)

Objetivos: fortalecer la seguridad durante el desarrollo y despliegue de aplicaciones mediante la asignación controlada de derechos administrativos y el escáner de elementos privados o de acceso restringido.

Implementación:

- Crear *Task* para el escáner.

La tarea está diseñada para escanear secretos en un directorio de código fuente utilizando la herramienta Trivy de Aqua Security que ayuda a identificar vulnerabilidades y problemas de seguridad en diferentes componentes.

Implementación en la etapa de entrega del *pipeline*

Para establecer un entorno seguro en *Tekton Pipelines* en la etapa de entrega se implementaron varios aspectos de distintos controles del *NIST SP 800-53*, en cumplimiento con el marco legal cubano que establece controles de seguridad para entidades estatales, que incluyen la gestión de accesos y protección de datos en la *Resolución 105/2020*.

Control AC-3 AC-6 AC-4 AC-10 de NIST SP 800-53

Implementación

- Crear un *namespace* y una cuenta de servicio con permisos mínimos.

Se crea una cuenta de servicio (*ServiceAccount*) dentro del denominado *namespace* previamente creado. La cuenta de servicio será

utilizada por los *pods* para acceder a los recursos necesarios dentro del clúster. Al establecer esta cuenta de servicio, se facilita la ejecución de tareas automatizadas y la implementación de procesos de integración continua.

- Aplicar en el clúster con *kubectl* apply -f <service-account.yaml>
- Definir un rol que solo permita ejecutar *pipelines* sin permisos administrativos.

Implementación en la etapa de despliegue del pipeline

En los entornos Kubernetes para el despliegue se utilizan múltiples tecnologías y herramientas (como Docker, Helm, Istio, entre otros) que pueden variar entre organizaciones. En esta etapa se proponen mejores prácticas de seguridad, según las características necesarias de cada proyecto o *software* desplegado en función de la organización. Recomendar estándares generales permite que cada equipo elija las herramientas más adecuadas para su contexto sin limitarse a una solución específica que puede no ser inclusiva para sus necesidades. La etapa de despliegue en un *pipeline* DevSecOps es crítica para garantizar la seguridad operacional, donde se aplicarán controles específicos del estándar *ISO/IEC 27001:2022*. Los controles seleccionados abarcan el A.12.1.3 (gestión de capacidad), A.8.23 (seguridad en servicios en red) y A.8.28 (protección en sistemas públicos).

Control A.12.3 del ISO/IEC 27001:2022 (Gestión de capacidad)

El control establece que el uso de los recursos debe monitorizarse, ajustarse y proyectarse para anticipar necesidades futuras, garantizando así el rendimiento requerido del sistema. Asimismo, enfatiza en la importancia de escanear y evaluar los componentes del sistema para identificar vulnerabilidades conocidas. Para implementarlo se recomienda añadir pasos en el *Task* de despliegue que cumplan funciones propuestas.

Objetivo: Limitar CPU por *pod* para evitar consumo excesivo.

Control A.8.23 del ISO/IEC 27001:2022 (Seguridad en red)

El control se enfoca en garantizar que las comunicaciones dentro de los sistemas estén protegidas contra accesos no autorizados, ataques y exposición de datos sensibles.

Objetivo: Restringir tráfico entre *pods* (solo *frontend* a *backend*).

Control A.8.28 del ISO/IEC 27001:2022 (Protección en sistemas públicos)

El control se refiere a la gestión de configuraciones y cambios en sistemas de información. Establece que las configuraciones de los sistemas deben ser documentadas, controladas y revisadas con regularidad para garantizar que no introduzcan vulnerabilidades. Además, los cambios en las configuraciones deben ser evaluados, probados y aprobados antes de su implementación para minimizar riesgos.

Objetivo: Asegurar comunicaciones con HTTPS (*Hyper Text Transfer Protocol Secure*), y evitar datos en claro. Forzar HTTPS en *Ingress* redirigiendo HTTP a HTTPS y usar HSTS.

Estrategia de validación del procedimiento propuesto

La validación de procedimientos es fundamental en el campo de la seguridad informática, ya que es un proceso riguroso que asegura que los métodos establecidos cumplan de forma efectiva con los objetivos propuestos. A través de este procedimiento, se verifica la capacidad de cada elemento, se identifican posibles fallos y se corrigen, demostrando su idoneidad para proteger los activos informáticos. Para esta investigación, se emplearon dos enfoques: simulación y métricas.

La Figura 2 detalla el procedimiento para desarrollar **la simulación en VirtualBox**. Esta guía describe cada etapa del proceso con el fin de asegurar la correcta aplicación de las técnicas y lograr una validación exitosa en un entorno real.

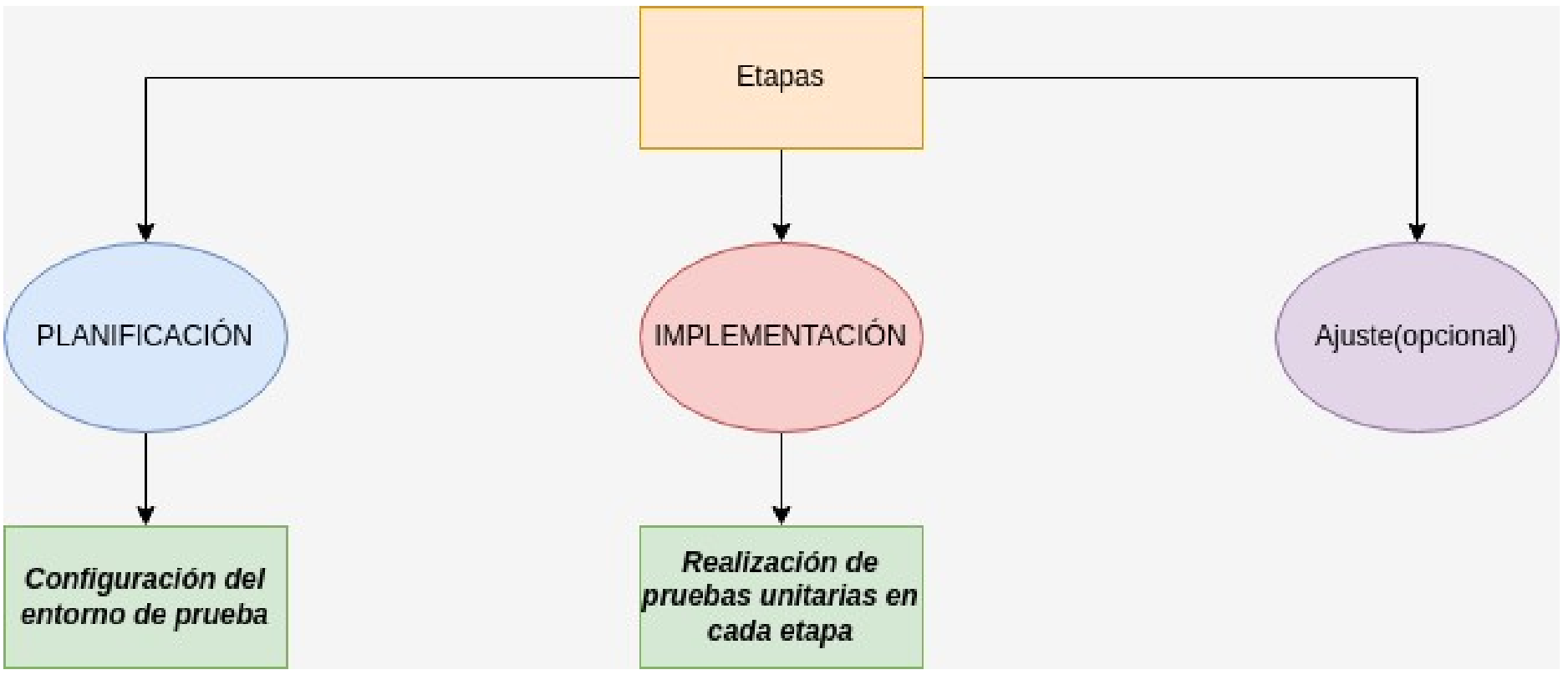

Figura 2. Etapas que componen el procedimiento empleado en la estrategia de validación simulación

Las dos etapas principales, junto con una fase opcional, facilitan el establecimiento de objetivos claros, la identificación de los recursos necesarios y la definición de criterios de éxito. La fase de ajuste incluye la corrección de hallazgos identificados, ajuste de configuraciones según resultados, identificación de áreas de mejora, documentación de cambios implementados y priorizar acciones correctivas según la necesidad empresarial.

Las fases materializan el procedimiento en acciones concretas. Esto facilita la reproducción controlada de escenarios de seguridad y la verificación práctica de los procedimientos. Además, contribuyen al análisis de los resultados obtenidos y generan conclusiones relevantes que sirven de base para recomendaciones orientadas a la mejora continua. La secuencia lógica de las etapas asegura que la simulación sea un proceso riguroso, en el que cada fase se apoya en la anterior para producir resultados confiables y aplicables.

El uso de **métricas** para determinar el cumplimiento normativo es fundamental para garantizar que la organización cumpla con los estándares de seguridad y regulaciones aplicables, lo que facilita la identificación de brechas y la mejora continua en las políticas de seguridad. Maverix es una solución clave para implementar DevSecOps debido a su capacidad de automatizar y centralizar métricas críticas de seguridad, integrando distintas herramientas en los *pipelines* CI/CD. Su modelo cuantificable permite priorizar vulnerabilidades, reducir riesgos y demostrar cumplimiento normativo, mientras fomenta una cultura de seguridad ágil (Maverix, 2021).

Maverix calcula el CP (cumplimiento de política) mediante la fórmula $CP = \frac{PC}{PT} \times 100$, donde PC son los puntos de control de seguridad implementados y PT el total de controles requeridos según estándares reconocidos. Esta métrica permite evaluar brechas de seguridad (ejemplo: un CP del 80 % indica un 20 % de controles pendientes) y priorizar su implementación para cumplir con normativas o mejorar la postura de seguridad. Se determina que el umbral de cumplimiento aceptable se establece cuando CP es mayor o igual al 75 %.

El umbral de cumplimiento del 75 % fue seleccionado como aceptable, en línea con recomendaciones de diferentes marcos internacionales:

• Según la sección 6.1.2 de la norma *ISO/IEC 27001*, los controles de seguridad pueden ajustarse según el nivel de riesgo.

• El marco del NIST CSF clasifica los niveles de madurez en cuatro categorías (Tier 1 a Tier 4), donde Tier 3 («Consistente») indica que la mayoría de los controles relevantes están en práctica. Un cumplimiento del 75 % puede corresponder con un nivel de madurez Tier 3, lo que significa que la organización tiene una postura de seguridad estable, pero aún no alcanza la excelencia, que sería Tier 4.

• Los *CIS Controls* (conjunto de mejores prácticas diseñadas para ayudar a las organizaciones a mejorar su ciberseguridad) priorizan los controles básicos y fundamentales. Para sistemas de baja criticidad, se recomienda implementar al menos el 75 % de los controles básicos.

Desde una perspectiva económica, alcanzar el 100 % de cumplimiento implicaría un incremento sustancial en los costos, sin ofrecer beneficios proporcionales en la reducción de riesgos. Por lo tanto, se concluye que un nivel de cumplimiento entre el 75 % y el 80 % es común en entornos similares (de prueba o de baja criticidad), lo que respalda la elección de este umbral como una opción razonable y práctica.

Conclusiones

Como resultado de la investigación realizada, se concluye que:

La comprensión de los conceptos fundamentales de DevSecOps es crucial para implementar prácticas de seguridad efectivas en el desarrollo de *software*. El análisis detallado de los estándares y mejores prácticas permite establecer un marco teórico-metodológico sólido que sustenta la implementación exitosa de procedimientos de seguridad, lo que garantiza la integridad y confiabilidad del proceso de desarrollo.

La implementación de estándares y mejores prácticas de seguridad en *pipelines* DevSecOps es fundamental para proteger la infraestructura empresarial y garantizar la integridad del *software*. El procedimiento

estructurado brindado permite identificar y mitigar vulnerabilidades para fortalecer la seguridad de la organización mediante un enfoque sistemático y controlado.

La validación de la propuesta consolida su credibilidad por medio de un enfoque reflexivo que permite ajustar los resultados según los datos obtenidos, lo que contribuye al avance del conocimiento científico. Las simulaciones efectuadas validaron que el flujo de trabajo es sólido y que el procedimiento alcanza los objetivos planteados, y esto asegura su aplicabilidad en la práctica y su capacidad para generar mejoras.

Integrar estándares y mejores prácticas de seguridad en el *pipeline* DevSecOps es fundamental para proteger proactivamente el *software* desde las etapas iniciales del desarrollo. Permite identificar y abordar vulnerabilidades antes de que se conviertan en problemas costosos, lo que asegura el cumplimiento normativo y mejorando la calidad del *software*.

Referencias bibliográficas

Antoniotti, L. (2021). *Tesi di Laurea Magistrale Pipeline DevSecOps* [Máster en Ingeniería Informática, POLITECNICO DI TORINO]. Electronico. http://webthesis.biblio.polito.it/id/eprint/20537

Atlassian. (2025, febrero 20). *What is SDLC? Software Development Life Cycle Explained*. Atlassian. https://www.atlassian.com/agile/software-development/sdlc

Chandramouli, R., Kautz, F., & Torres-Arias, S. (2024). *Strategies for the integration of software supply chain security in DevSecOps CICD pipelines*: (No. NIST SP 800-204D; Número NIST SP 800-204D, p. NIST SP 800-204D). National Institute of Standards and Technology (U.S.). https://doi.org/10.6028/NIST.SP.800-204D

GitLab 2024 *Global DevSecOps Report* | GitLab. (2025, febrero 26). Nuxt-App. https://about.gitlab.com/developer-survey/

ISO/IEC27001:2022.(2022).ISO.https://www.iso.org/standard/27001

Kubernetes. (2024). *Overview*. Kubernetes. https://kubernetes.io/docs/concepts/overview/

Madnick, P. S. E. (2023). *The Continued Threat to Personal Data: Key Factors Behind the 2023 Increase*.

Maverix. (2021). DevSecOps automation platform: Metrics-driven security compliance. Maverix Security Solutions. https://www.maverix.com/resources/devsecops-automation-whitepaper

Muñoz, C. D. (2023). DevOps en el desarrollo de software: Integración y entrega continua. *Polo del Conocimiento: Revista científico-profesional*, *8*(9 (SEPTIEMBRE 2023)), 603-615.

Saber, H. (2018). *Consejo de Estado*.

Tekton. (2023, noviembre 13). *How it all started The Tekton Story—Tekton*. https://blog.tektonlabs.com/tekton-labs-blog/how-it-all-started-the-tekton-story/

Wilbur L. Ross, Jr., Secretary. (2020). *Security and Privacy Controls for Information Systems and Organizations*.

Xygeni. (2024). *Mejores prácticas de CI/CD: Transformando el desarrollo de software*| xygeni. https://xygeni.io/es/blog/cicd-best-practices-transforming-software-development/

Zayni, M., & Nassereddine, B. (2020). (PDF) *DevSecOps Practices for an agile and secure it service management*. https://www.researchgate.net/publication/338659928_DevSecOps_PRACTICES_FOR_AN_AGILE_AND_SECURE_IT_SERVICE_MANAGEMENT

Zhu, B., & Zou, E. (2021). *Design of intelligent bedroom system based on pure off-line speech recognition technology*. https://ieeexplore.ieee.org/document/9407425/references#references

INVESTIGACIÓN

Despliegue de Honeypots para la detección proactiva de amenazas cibernéticas

Deployment of honeypots for proactive detection of cyber threats

Ing. Heidy Rodríguez Malvarez*, Ing. Elizabeth Molina Mena[2], Dra.C. Mónica Peña Casanova[3]

Recibido: 11/2025 | Aceptado: 11/2025 | Publicado: 11/2025

Resumen

En un entorno digital cada vez más amenazado por actores maliciosos, la necesidad de implementar mecanismos proactivos de seguridad se ha vuelto crítica. Este artículo presenta el diseño, la configuración y la validación de un sistema de honeypots, con especial énfasis en la herramienta T-Pot, orientado a la detección proactiva de amenazas cibernéticas. T-Pot, reconocido por su arquitectura modular y versatilidad, se despliega como el núcleo de una solución capaz de capturar, analizar y monitorear actividades malintencionadas en tiempo real. En primer lugar, se aborda el estado de la cuestión de los honeypots y sus aplicaciones en la ciberseguridad, destacando su relevancia como técnica de engaño para observar y analizar el comportamiento de los atacantes. En segundo lugar, se define una arquitectura flexible y escalable implementada en un entorno de laboratorio controlado, donde se simulan ataques

* Departamento de Infraestructuras Tecnológicas, Facultad Ciberseguridad, Universidad de las Ciencias Informáticas. heidyrm@uci.com

[2] Dirección de Seguridad Informática, Universidad de las Ciencias Informáticas. angelagv@uci.cu

[3] Departamento de Infraestructuras Tecnológicas, Decana Facultad de Ciberseguridad, Universidad de las Ciencias Informáticas. monica@uci.cu

diversos para evaluar la eficacia del sistema. Los resultados obtenidos permiten validar la capacidad del honeypot para identificar intentos de intrusión y comportamientos sospechosos lo que demuestra la efectividad de T-Pot como herramienta para la detección proactiva de amenazas.

Palabras clave: amenazas cibernéticas, ciberseguridad, detección proactiva, honeypot, T-Pot

Abstract

In an increasingly threatened digital environment, the need to implement proactive security mechanisms has become critical. This article presents the design, configuration and validation of a honeypot system, focusing on the T-Pot tool, aimed at proactive cyber threat detection. T-Pot, recognized for its modular architecture and versatility, is deployed as the core of a solution capable of capturing, analyzing, and monitoring malicious activities in real time. The work begins with an overview of honeypots and their applications in cybersecurity, emphasizing their relevance as a deception technique to observe and analyze attacker behavior. A flexible and scalable architecture is defined and implemented in a controlled laboratory environment where various attacks are simulated to evaluate system effectiveness. The results validate the honeypot's ability to identify intrusion attempts and suspicious behavior, demonstrating T-Pot's effectiveness as a tool for proactive threat detection.

Keywords: *Cyber Threats, Cybersecurity, Proactive Detection, Honeypot, T-Pot*

Introducción

En el panorama contemporáneo de la ciberseguridad, las organizaciones enfrentan una creciente sofisticación y diversificación de las amenazas cibernéticas (Fadziso et al., 2023). Los actores maliciosos, que abarcan desde individuos aislados hasta grupos organizados, emplean tácticas cada vez más avanzadas con el objetivo de comprometer sistemas informáticos, sustraer información confidencial o interrumpir servicios esenciales (Cherqi et al., 2023). Esta

situación ha impulsado la necesidad de adoptar mecanismos proactivos que permitan anticipar los ataques y responder de manera eficaz antes de que estos causen daños irreversibles.

De forma tradicional, las estrategias de defensa de los centros de operaciones de seguridad (SOC) se han basado en el uso de herramientas como los sistemas de detección y prevención de intrusiones (IDS/IPS) y los programas antivirus. Sin embargo, estos métodos presentan limitaciones ante la evolución de las técnicas de evasión utilizadas por los atacantes, quienes logran eludir las soluciones que dependen de firmas o patrones antes conocidos (Skopik et al., 2016). De ahí que resulte imperativo implementar soluciones que trasciendan la detección reactiva y que permitan observar y analizar el comportamiento real de los atacantes en un entorno controlado.

En este contexto, los honeypots emergen como una tecnología eficaz para la localización y el estudio de actividades maliciosas. Estos sistemas, diseñados para parecer vulnerables y atractivos a los atacantes, actúan como señuelos y permiten registrar sus tácticas, técnicas y procedimientos sin poner en riesgo los activos reales de la organización (Lorusso Montiel & Uc Ríos, 2022). La información obtenida mediante honeypots resulta invaluable para fortalecer las defensas de la infraestructura tecnológica y para anticiparse a futuras amenazas.

El presente trabajo tiene como objetivo evaluar el despliegue de honeypots como medida proactiva para la detección temprana de amenazas cibernéticas en entornos organizacionales, con especial énfasis en el uso de la herramienta T-Pot reconocida por su arquitectura modular, capacidad de integración y soporte para múltiples servicios de emulación.

Para alcanzar este propósito, se implementó una solución basada en un entorno de laboratorio controlado que simula diferentes tipos de ataques y analizar la efectividad de la herramienta en la captura, el monitoreo y la correlación de datos relacionados con incidentes de seguridad. Este enfoque contribuye no solo a la validación experimental del honeypot, sino también al fortalecimiento de las estrategias de defensa cibernética de las organizaciones.

Materiales y métodos

El desarrollo de esta investigación se sustentó en un enfoque metodológico mixto. Se combinaron métodos teóricos y empíricos para diseñar, implementar y validar una arquitectura de honeypots orientada a la detección proactiva de amenazas cibernéticas en entornos organizacionales.

Diseño general de la investigación

La investigación se estructuró en tres fases fundamentales: diseño, implementación y validación del honeypot T-Pot.

En la primera fase se realizó una revisión bibliográfica sobre los mecanismos de detección proactiva de amenazas y los diferentes tipos de honeypots, esto posibilitó la selección de la herramienta más adecuada para los requerimientos de la Empresa de Tecnologías de la Información (ETI) de BioCubaFarma.

En la segunda fase se diseñó una arquitectura experimental basada en el despliegue de T-Pot en un entorno controlado de laboratorio, donde se simularon ataques de red y web, para reproducir condiciones similares a las que enfrenta un centro de operaciones de seguridad (COS).

En la tercera fase se realizó la validación empírica del sistema, a través el uso de las métricas de detección, correlación de eventos y análisis de alertas. Esta estructura metodológica demostró la eficacia de T-Pot como mecanismo complementario a las soluciones tradicionales de seguridad (IDS/IPS y antivirus) utilizadas por la ETI (Empresa de Tecnologías de la Información, 2024).

Entorno y herramientas tecnológicas utilizadas

El entorno experimental se desplegó en una infraestructura virtual compuesta por:

- Servidor principal (Ubuntu Server 22.04 LTS): se implementó la plataforma T-Pot, una solución de honeypot de código abierto desarrollada por Deutsche Telekom.
- Máquinas virtuales de ataque: se utilizaron para simular amenazas con herramientas como Nmap y OWASP ZAP, desde un sistema operativo Kali Linux.

- Red virtual controlada (VirtualBox): se configura para aislar los experimentos del entorno productivo y garantizar la seguridad del laboratorio.

En la Figura 1 se muestra el entorno explicado anteriormente:

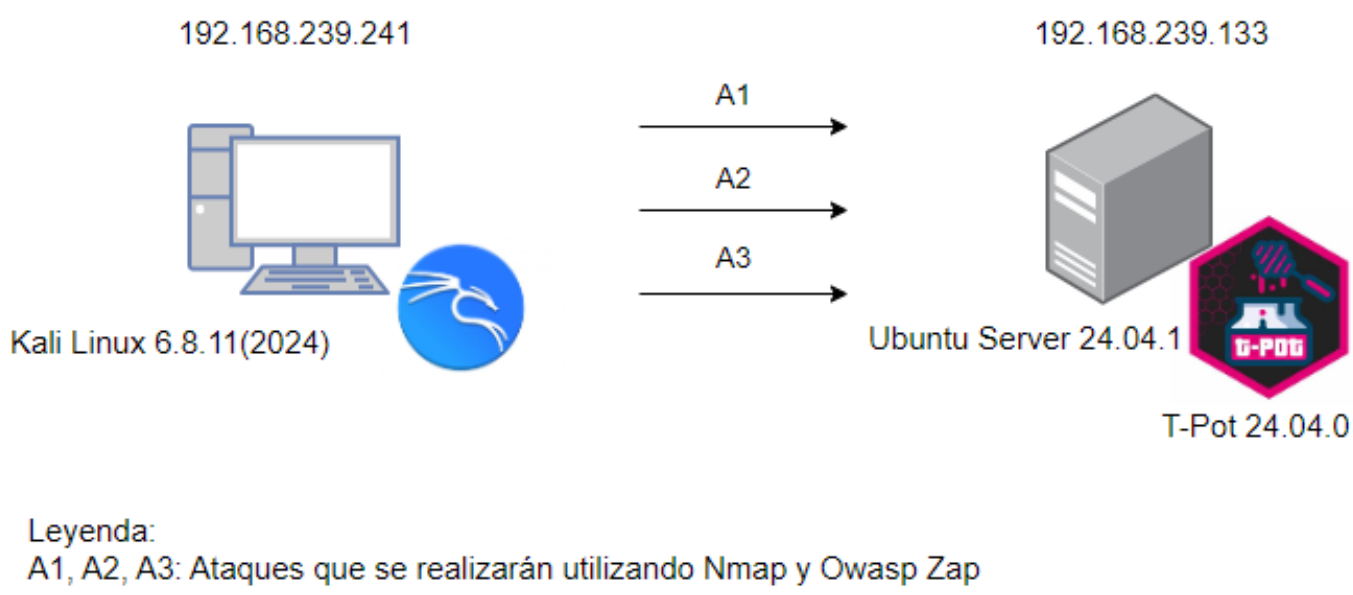

Figura 1. Entorno de laboratorio controlado

El honeypot T-Pot integra diversos módulos de emulación de servicios, entre que destacan Cowrie (SSH/Telnet), Dionaea *(malware catcher)*, *E*lasticpot *(Elasticsearch* honeypot), Heralding (servidor de autenticación) y Mailoney (SMTP falso). Esta arquitectura modular facilita la recolección y clasificación de eventos de acuerdo con el vector de ataque y el servicio objetivo (Telekom Security, 2024).

Para el monitoreo y análisis de datos se emplearon herramientas complementarias como:

- Kibana: permite la visualización de métricas e indicadores de ataque.
- Elasticsearch: funciona como motor central de almacenamiento y análisis.
- Suricata: se emplea para la detección y correlación de tráfico malicioso.
- Spiderfoot y Elasticvue: se utiliza para el reconocimiento automatizado de *hosts* y consultas a los índices de datos capturados.

Métodos científicos

Se aplicaron los siguientes métodos teóricos y empíricos:

- Histórico-lógico: aplicado para analizar la evolución de las amenazas cibernéticas y la progresión de las estrategias de defensa digital.

- Analítico-sintético: para estudiar las tecnologías de honeypots existentes, identificar sus ventajas y desventajas, y fundamentar la elección de T-Pot como solución experimental.
- Modelación: para diseñar el entorno de red simulado que replicó la infraestructura tecnológica de la ETI, con el fin de evaluar la efectividad de la herramienta bajo condiciones reales de ataque.
- Medición: utilizada para valorar el desempeño del honeypot en función de la cantidad y tipo de eventos detectados, frecuencia de alertas generadas y efectividad en la correlación de datos.

Estas herramientas y métodos permitieron establecer un marco experimental sólido, caracterizado por una configuración controlada del honeypot que garantizó la reproducibilidad del estudio y la validez de los resultados obtenidos.

Resultados y discusión

Durante la fase de validación, el honeypot T-Pot demostró su capacidad para capturar y analizar una amplia gama de actividades maliciosas en tiempo real. El sistema registró eventos relacionados con escaneos de red, intentos de autenticación no autorizada, ataques de denegación de servicio y explotación de vulnerabilidades en servicios simulados.

En la Figura 2 muestra un entorno muy activo en términos de ciberamenazas, con un total de 34,818 ataques hacia múltiples honeypots. De este total de ataques se puede establecer que el *Honeytrap* fue el honeypots que más ataques detectó con un total de 32,507 ataques, lo que indica la prevalencia de escaneos automatizados y conexiones no solicitadas dirigidas a servicios frecuentes.

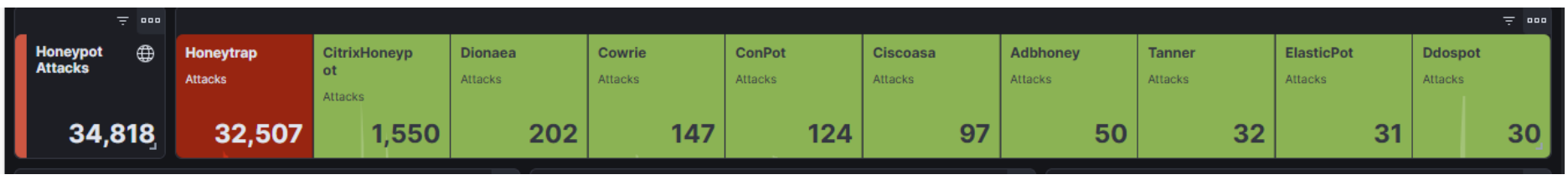

Figura 2. Ataques detectados por el T-Pot

En las pruebas realizadas, se observó que los servicios más atacados fueron SSH (puerto 22), HTTP (puerto 80) y SMB (puerto 445), lo que evidencia la prevalencia de tácticas automatizadas de

reconocimiento y explotación. Estas actividades fueron visualizadas mediante paneles de Kibana, donde se generaron histogramas, gráficos de densidad y mapas geográficos que representaron la procedencia y volumen de los ataques.

El motor Suricata generó múltiples alertas clasificadas según la severidad y tipo de incidente como se presenta en la Figura 3. La categoría predominante es *Generic Protocol Command Decode*, representada en color rojo, que muestra un volumen muy alto de actividad sospechosa relacionada con comandos genéricos de protocolos, asociados a intentos de reconocimiento o escaneo masivo. En menor proporción, se registran otras categorías como *Potentially Bad Traffic* (color verde), que incluye tráfico con características anómalas; *Web Application Attack* (color azul), que representa intentos de ataques dirigidos a aplicaciones web; y *A Network Trojan was Detected* (color morado), relacionada con patrones de troyanos. Además, se identificó tráfico marcado como *Not Suspicious Traffic* (color rosado claro), el cual, aunque no presentó de forma clara comportamientos maliciosos, fue registrado para análisis.

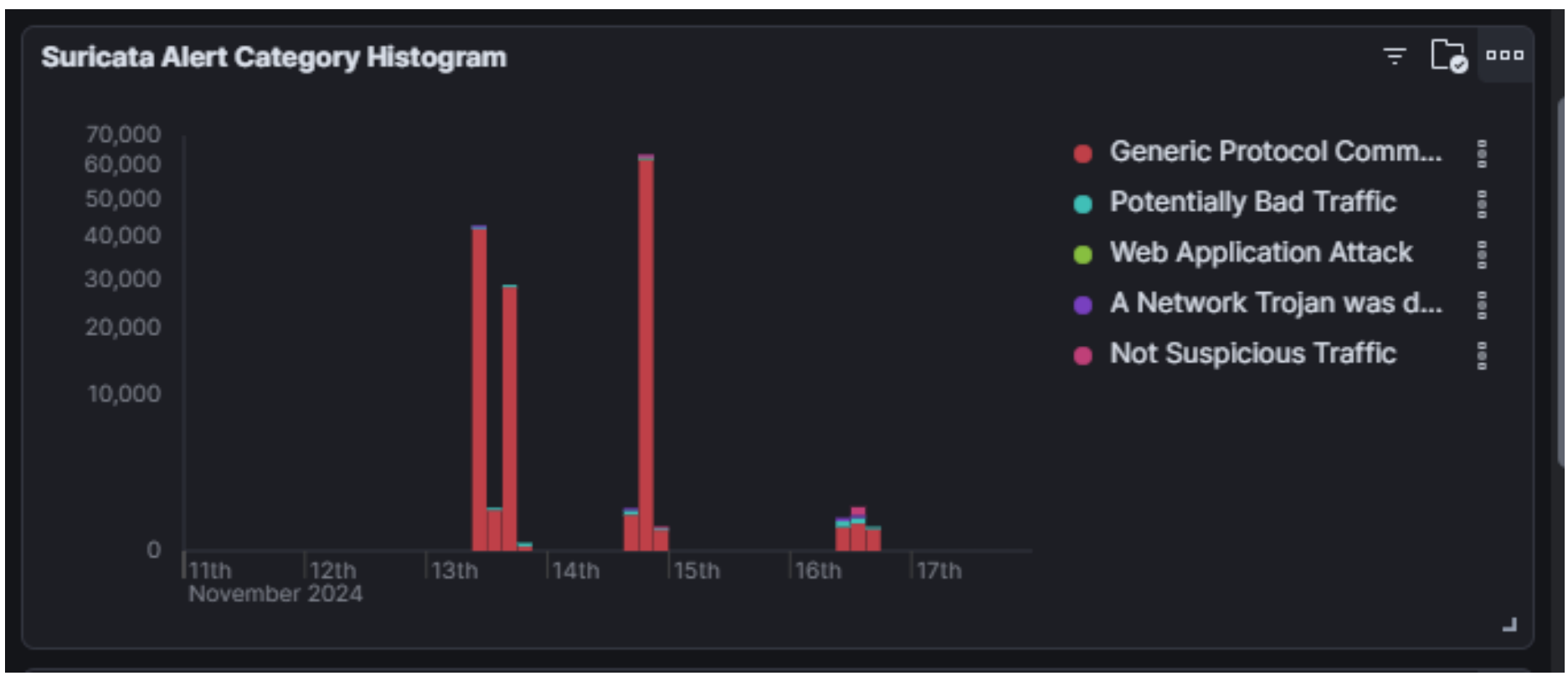

Figura 3. Alertas generadas por Suricata

Uno de los hallazgos más relevantes fue la identificación de patrones repetitivos en las tácticas de ataque. La Figura 4 presenta nubes de palabras (*tag clouds*) que visualizan información relacionada con los intentos de autenticación en el tráfico capturado por el honeypot. En la nube de nombres de usuario se observan varias

entradas de texto relacionadas con los nombres de usuario utilizados en las solicitudes. La palabra «anonymous» es la más destacada, lo que indica que muchos intentos de acceso se realizaron con un nombre de usuario anónimo o no identificado. Otras entradas, como «sip» o «From: sip:nm@nm», podrían sugerir intentos de acceso a servicios relacionados con el protocolo SIP (*Session Initiation Protocol*), es muy común su utilización en comunicaciones VoIP. Otros términos como «Max-Forwards: 70» y «Call-ID: 50000» refuerzan la hipótesis de un ataque o escaneo dirigido a servicios VoIP. Por su parte, la nube de contraseñas resalta los intentos de contraseñas observados durante los ataques. Algunas de las contraseñas más comunes fueron «anonymous», «(empty)», «IEUser@», y «lolo», lo que indica que los atacantes intentaron acceder utilizando credenciales débiles o fácilmente adivinables. El término «(empty)» implica que se intentaron iniciar sesiones sin contraseña, mientras que «anonymous» refuerza el uso de accesos no autenticados.

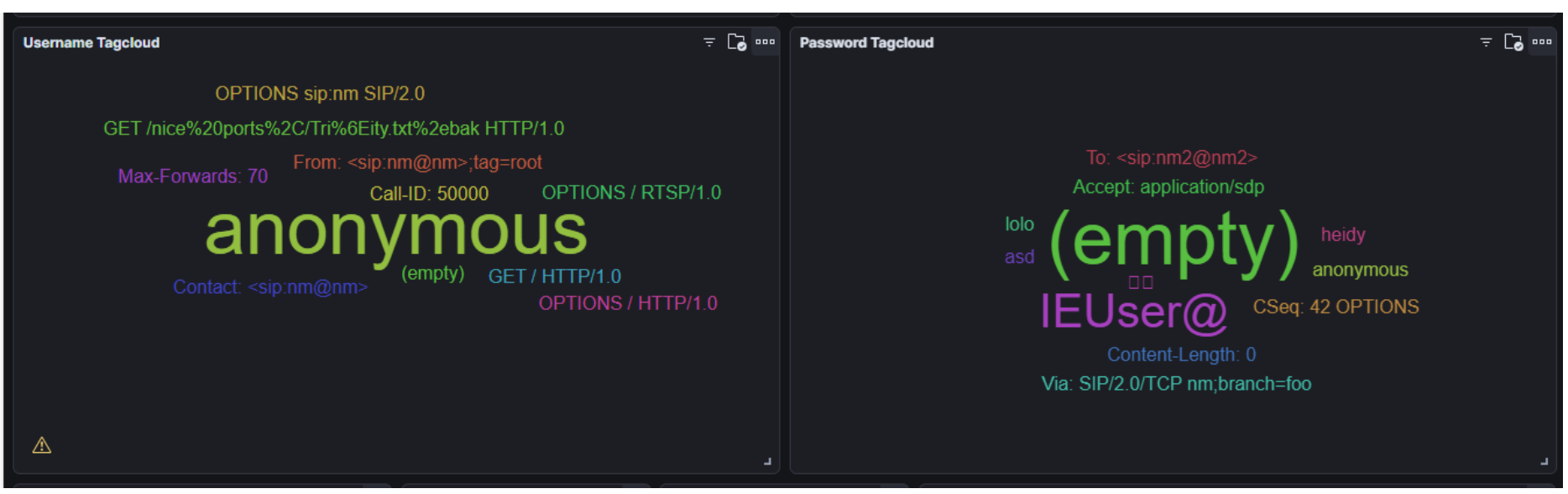

Figura 4. Nube de palabras

Los resultados obtenidos demuestran que el despliegue de T-Pot no solo es útil para la detección temprana de amenazas, sino que también contribuye a la inteligencia de amenazas cibernéticas (CTI), lo que permite a las organizaciones observar el ciclo completo de ataque, desde la fase de exploración hasta la ejecución.

Además, el uso de herramientas integradas como *Elasticvue* facilitó las consultas precisas sobre los datos recolectados, esto optimizó el análisis forense y la clasificación de incidentes. Esta integración pone de manifiesto la fortaleza de T-Pot como plataforma escalable y adaptable, capaz de incorporarse en infraestructuras corporativas o académicas para fines de monitoreo, formación y análisis de amenazas.

Conclusiones

El despliegue del sistema T-Pot como honeypot avanzado en un entorno de laboratorio controlado validó su efectividad como mecanismo proactivo de detección de amenazas cibernéticas. Los resultados confirman que esta herramienta es capaz de identificar intentos de intrusión, escaneos de red y patrones de comportamiento malicioso, ofreciendo una visión integral del panorama de amenazas.

La arquitectura modular de T-Pot, junto con su integración nativa con herramientas de análisis como Kibana, Elasticsearch y Suricata, proporciona un entorno eficiente para la recopilación, la visualización y la correlación de eventos en tiempo real. Esta capacidad constituye una ventaja significativa frente a los mecanismos tradicionales, al permitir una respuesta anticipada ante amenazas emergentes.

Además, el estudio evidencia que la adopción de honeypots puede contribuir a fortalecer las estrategias defensivas del centro de operaciones de seguridad de la ETI, al servir como fuente de información para la inteligencia de amenazas, la capacitación del personal técnico y la investigación de nuevas tendencias de ataque. Como líneas futuras, se propone integrar T-Pot con sistemas SIEM (como OSSIM o ELK), para mejorar la correlación automática de eventos y explorar el despliegue de honeypots distribuidos que posibiliten ampliar el espectro de detección y aumentar la cobertura de la red.

En síntesis, el trabajo realizado demuestra que la implementación de honeypots constituye una alternativa viable, escalable y de bajo costo para el fortalecimiento de la ciberseguridad en infraestructuras críticas y entornos empresariales cubanos.

Referencias bibliográficas

Cherqi, O., El Omri, A., & Achahbar, A. (2023). *Cyber Threat Intelligence approaches for proactive security*. *Journal of Cyber Defense Studies*, 9(2), 45–59.

Empresa de Tecnologías de la Información. (2024). *Centro de Operaciones de Seguridad de BioCubaFarma*. La Habana, Cuba.

Fadziso, F., Moyo, T., & Zhou, M. (2023). *The evolution of cyber threats and defense mechanisms. International Journal of Information Security*, 22(3), 214–229.

Lorusso Montiel, J., & Uc Ríos, A. (2022). *Honeypots como estrategia de defensa cibernética. Revista de Seguridad Informática*, 18(2), 45–53.

Skopik, F., Settanni, G., & Fiedler, R. (2016). *A survey on threat intelligence sharing platforms. Computers & Security*, 61, 1–16.

Telekom Security. (2024). *T-Pot Documentation: Multi-Honeypot Platform*. Deutsche Telekom AG. Recuperado de https://github.com/telekom-security/tpotce

INVESTIGACIÓN

Capacitación Proactiva contra correos phishing mediante simulaciones con la herramienta GoPhish

Proactive training against phishing emails using simulations with the GoPhish tool

Ing. Rubén Olabarrieta Rivera[*], Ing. Yuneisy Quintero Prieto[2], Ing. Raúl Sanchez Arañó[3]

Recibido: 11/2025 | Aceptado: 11/2025 | Publicado: 12/2025

Resumen

El *phishing*, una amenaza centrada en la explotación del factor humano, representa un alto riesgo para las organizaciones. Frente a la insuficiencia de las capacitaciones teóricas tradicionales, este estudio implementó un programa de simulaciones controladas que utiliza la herramienta de código abierto GoPhish. A lo largo de cuatro campañas sucesivas que emplearon distintos vectores de ataque, se observó una tendencia decreciente en las tasas de interacción y una reducción significativa en la entrega de credenciales. El análisis de los usuarios participantes reveló una clara curva de aprendizaje, donde la mayoría mostró mejoras sustanciales en su capacidad de detección. Al mismo tiempo, se registró un aumento considerable en los reportes voluntarios de correos sospechosos en el entorno real, de manera especial en áreas de alta exposición. Los resultados demuestran que la combinación de entrenamiento práctico, retroalimentación inmediata y simulaciones periódicas transforma de forma efectiva al usuario

[*] TC Mariel. Cuba. rubenor@tcmariel.cu

[2] TC Mariel. Cuba. yuneisyqp@tcmariel.cu

[3] TC Mariel. Cuba. raulsa@tcmariel.cu

de un eslabón vulnerable en una barrera de defensa proactiva, que fortalece de manera sostenible la cultura de ciberseguridad organizacional.

Palabras clave: phishing, GoPhish, simulación de ciberseguridad, entrenamiento práctico

Abstract

Phishing, a threat exploiting the human factor, poses a critical risk to organizations. Confronting the inadequacy of traditional theoretical training, this study implemented a controlled simulation program using the open-source tool GoPhish. Across four successive campaigns employing distinct attack vectors, a clear decreasing trend was observed in interaction rates and credential submission was significantly reduced. Analysis of participating users revealed a strong learning curve, with most showing substantial improvements in detection capabilities. Concurrently, a considerable increase in voluntary reporting of real suspicious emails was recorded, particularly in high-exposure areas. The findings confirm that combining practical training, immediate feedback, and periodic simulations effectively transforms users from a vulnerable link into a proactive defense barrier, thereby sustainably strengthening the organizational cybersecurity culture.

Keywords: *phishing, GoPhish, cybersecurity simulations, practical training*

Introducción

El panorama de la ciberseguridad actual se caracteriza por una evolución constante y sofisticada de las amenazas, donde el correo electrónico permanece como uno de los vectores de ataque más prolíficos y efectivos. Entre estas amenazas, el *phishing* es definido como una forma de fraude en línea cuyo objetivo es robar información confidencial mediante el envío de comunicaciones que suplantan la identidad de una entidad legítima (Workman, 2008). Este se destaca no por su complejidad técnica, sino por su ingeniosa explotación de un eslabón vulnerable: el factor humano. Los atacantes han perfeccionado sus tácticas para crear mensajes fraudulentos tan convincentes

que vulneran con facilidad las defensas perimetrales mejor configuradas, dirigiéndose directamente a la psicología de los empleados.

Las consecuencias de un ataque de *phishing* exitoso son severas, y van desde el compromiso de credenciales y el robo de información sensible hasta infecciones con *ransomware* que pueden paralizar las operaciones completas de una empresa, esto genera cuantiosas pérdidas financieras y daños irreparables a la reputación corporativa. Frente a este escenario, la sola implementación de soluciones técnicas defensivas (como filtros de correo, antivirus y *firewall*) se ha revelado como insuficiente.

La defensa más crítica y subestimada radica en la capacitación continua y consciente de los usuarios finales. Para lograr esto, las empresas deben adoptar un enfoque proactivo que trascienda las capacitaciones tradicionales teóricas, mediante la simulación de ataques controlados que permitan medir y elevar el nivel de preparación real de su capital humano. En este contexto, herramientas de simulación de correos *phishing* como GoPhish emergen como recursos invaluables, lo que permite a los especialistas de ciberseguridad, ejecutar y analizar campañas de *phishing* simuladas internamente en las empresas. Esta práctica no solo identifica a los empleados más vulnerables para brindarles capacitación dirigida, sino que también sienta las bases para una cultura organizacional de ciberseguridad robusta y resistente, para transformar al eslabón más débil en una primera línea de defensa efectiva.

El objetivo principal de este trabajo es medir la susceptibilidad de la fuerza laboral, evaluar la efectividad de las herramientas de simulación para el entrenamiento proactivo en la detección de correos *phishing* e identificar áreas de mejora.

Materiales y métodos

Contexto de estudio y justificación del enfoque

El estudio se realizó en una empresa de servicios logísticos con más de 200 empleados que usan a diario correo electrónico y más del 50 % mantiene algún tipo de intercambio con clientes y proveedores internacionales lo que aumenta de manera exponencial la interacción

con correos electrónicos tipo *phishing* reales y su conocimiento de identificación es vital en la detección y minimización de incidentes. Para enfrentar esta amenaza, se implementó un programa de simulaciones controladas.

El *phishing* no es una amenaza nueva, pero su prevalencia y complejidad continúan en un crecimiento alarmante. Los reportes anuales de actores líderes en la industria proveen datos alarmantes que ilustran la magnitud del problema. El Verizon *2025* Data Breach Investigations Report (DBIR) consolida esta tendencia, situando a este patrón entre los tres principales desde 2019 (Verizon, 2025). Este informe destaca que la sofisticación ha evolucionado de forma significativa; los ataques ya no se basan en propuestas increíbles de trabajo o premios, sino en campañas prolongadas donde los atacantes construyen relaciones de confianza con las víctimas durante un periodo de tiempo.

La efectividad de esta técnica radica en su ingeniería social, un método que «influye psicológicamente en una persona para que realice una acción o revele información confidencial» (ENISA, 2022). Los actores de amenazas investigan de forma minuciosa a sus víctimas, a través de la creación de correos muy personalizados (*spear phishing*) que imitan de manera perfecta comunicaciones legítimas de superiores, colegas o servicios de confianza. El DBIR de Verizon (2025) corrobora que en la actualidad se debe ser cauteloso incluso ante los mensajes que parecen proceder de clientes o proveedores establecidos, lo que hace muy difícil su detección incluso para usuarios experimentados.

El usuario, el “eslabón débil”

La referencia tradicional en ciberseguridad ha caracterizado al usuario final como “el eslabón más débil” de la cadena de seguridad. Esta perspectiva, sin embargo, es contraproducente y se reevalua. Investigaciones en el campo de la seguridad de la información sugieren que es más efectivo ver al empleado no como un problema a ser mitigado, sino como un “factor humano” que debe ser empoderado y capacitado (Hadlington, 2017), «los empleados no son solo usuarios, sino guardianes activos de la ciberseguridad» (Hernández, 2025). Los

usuarios no fallan por negligencia inherente, sino porque la capacitación que reciben a menudo es genérica, esporádica y desconectada de las tácticas reales utilizadas por los atacantes en la actualidad.

Un estudio publicado en *Computers & Security* encontró que la formación tradicional basada en concienciación tiene un efecto limitado en el comportamiento a menos que se complemente con experiencias prácticas y retroalimentación inmediata (Pattinson et al., 2015). Es aquí donde entrenar con la realidad se vuelve crucial, la confianza cero no solo se trata de limitar accesos, sino de empoderar a las personas para que tomen decisiones informadas (Hernández, 2025). Al exponer a los usuarios a simulaciones de *phishing* realistas, se activa el aprendizaje experiencial, permitiéndoles reconocer las señales de un ataque (como remitentes extraños, URL sospechosas o urgencia injustificada) en un entorno seguro donde un error se convierte en una lección y aprendizaje, no en un desastre.

Herramienta seleccionada: GoPhish

Para implementar un programa de entrenamiento basado en simulaciones, las empresas necesitan herramientas accesibles y efectivas que evalúen la susceptibilidad de los empleados a través de ataques controlados. Si bien el mercado está dominado por soluciones comerciales reconocidas, tales como KnowBe4 y otras, líderes identificadas en el informe “Forrester Wave: Security Awareness and Training Solutions, Q1 2022” (Budge, et al., 2022), la elección de una plataforma depende a menudo del presupuesto y las necesidades específicas. Para organizaciones que priorizan la flexibilidad y la ausencia de costes de licencia, GoPhish surge como una alternativa de código abierto con una amplia aceptación.

GoPhish es una plataforma que permite a los profesionales de seguridad:

- Diseñar campañas de *phishing* altamente personalizables que replican las tácticas actuales.
- Dirigir estas campañas a grupos específicos de empleados de manera controlada.
- Medir con precisión métricas clave como la tasa de clic y la tasa de entrega de credenciales.

El valor de GoPhish, respaldado por una comunidad activa y documentación extensa, radica en la capacidad para cerrar el ciclo de aprendizaje: de la teoría a la práctica, y de la práctica a la mejora continua.

Diseño de las campañas y procedimiento experimental

Para las simulaciones se utilizaron dos vectores de ataque diferentes:

- Tipo 1: Envío de correo de solicitud de actualización de credenciales en un servicio web.
- Tipo 2: Envío de correo de solicitud de descarga de fichero.

Se ejecutaron cuatro campañas simuladas cronológicas a lo largo de dos años (2024-2025) (Tabla 1):

- Campaña 1 (Tipo 1). Actualización de credenciales en el *Webmail* de la empresa.
- Campaña 2 (Tipo 2). Solicitud de descarga de fichero «Actualización de políticas de seguridad de la empresa».
- Campaña 3 (Tipo 2). Solicitud de descarga de fichero «Listado de productos de feria».
- Campaña 4 (Tipo 1). Actualización de credenciales en el ERP (*Enterprise Resource Planning*) empresarial.

Tabla 1. Campañas en orden cronológico

	Usuarios	Clic en el enlace	Entregaron credenciales	Descarga de fichero	Reporte de usuarios
Campaña 1	176	21	10	--	0
Campaña 2	173	17	--	17	8
Campaña 3	227	40	--	40	8
Campaña 4	132	16	0	--	11

Medidas correctivas y recolección de datos

Después de cada Campaña, y en especial después de los resultados iniciales, se implementaron acciones correctivas. Tras la Campaña 1 se elaboró un Plan de Capacitación específico sobre identificación de

correos *phishing* orientado a las personas que no pasaron las simulaciones con materiales detallados y ejercicios prácticos de identificación. Además, se prepararon capacitaciones generales para todos los trabajadores sobre correos *phishing* y sesiones de concienciación general con los usuarios, enfocadas en la importancia de notificar de manera proactiva cualquier comunicación sospechosa al equipo de ciberseguridad, con el fin de agilizar la respuesta ante incidentes reales. Los datos se recopilaron de forma automatizada por la herramienta GoPhish (clic, entregas de credenciales) y con reportes manuales de los usuarios a la Dirección de IT.

Resultados y Discusión

Resultados por campaña y análisis de la evolución

La Campaña 1 consistió en la simulación de un correo electrónico que, a través de un enlace fraudulento, solicitaba a los usuarios la actualización de sus credenciales (Figura 1) en un portal que emulaba la interfaz del Webmail corporativo (Figura 2). Tras la introducción de los datos por parte del usuario, la solicitud era redirigida al portal oficial de la empresa, registrándose el evento como un intento de *phishing* exitoso. Los resultados de esta campaña revelaron una tasa de clic del 20 % y, de estos, una tasa de entrega de credenciales del 48 %. Estas cifras, alarmantes al inicio, evidenciaron una clara vulnerabilidad del personal ante este tipo de ataques, a pesar de la existencia de capacitaciones periódicas en la empresa. Un hallazgo adicional crítico fue la casi nula notificación de los correos fraudulentos por parte de los empleados.

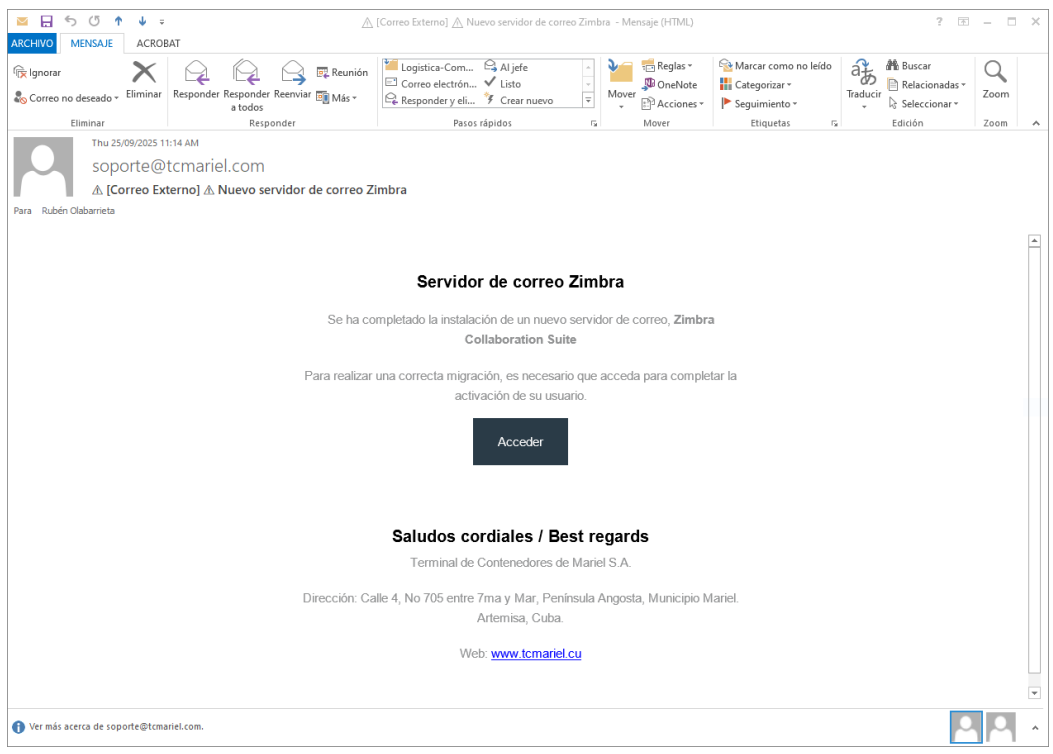

Figura 1. Correo de solicitud de actualización de credenciales en el Webmail de la empresa

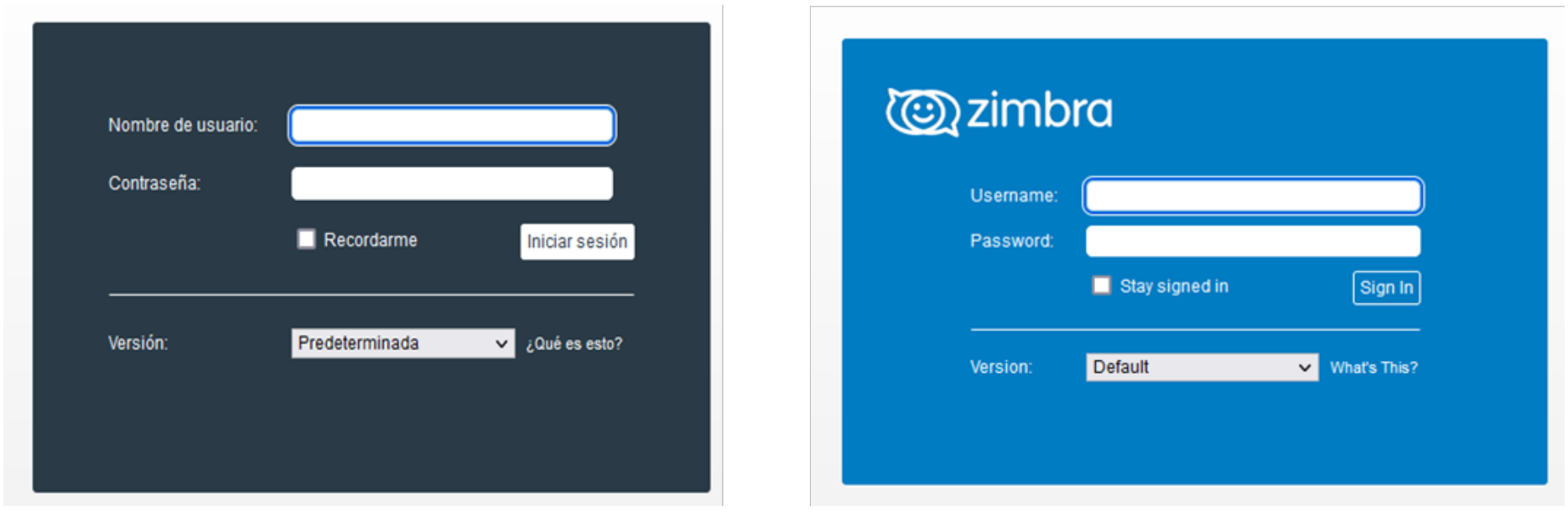

Figura 2. Interfaz falsa (izquierda) y real (derecha) del Webmail corporativo

La Campaña 2 se diseñó para evaluar la efectividad de las medidas de concientización implementadas, con el uso de un vector de ataque distinto: la descarga de un archivo malicioso. En esta simulación, el correo electrónico instaba a los usuarios a descargar un archivo bajo la denominación «Actualización de políticas de seguridad de la empresa». El evento se registró como un intento de *phishing* exitoso si el usuario procedía a descargar el fichero. Los resultados mostraron una mejora sustancial: la tasa de clic y descarga del archivo se redujo al 10 %. De manera aún más significativa, se registró un incremento en las notificaciones por parte de los empleados, pasando de un 0 % en la campaña inicial a un 4 %. Este aumento en los reportes, aunque modesto, sugiere una mayor concientización y una respuesta proactiva del personal, lo que evidencia la efectividad preliminar de las capacitaciones específicas y generales impartidas con posterioridad a la primera simulación.

La Campaña 3 incrementó de manera significativa el nivel de sofisticación mediante la aplicación de técnicas avanzadas de ingeniería social. El correo, diseñado con un mayor apego a la realidad operativa de la empresa, solicitaba la descarga de un archivo denominado «Listado de productos de feria». Al ejecutarse, este archivo iniciaba una secuencia automatizada que simulaba comportamientos maliciosos, como la ejecución consecutiva de aplicaciones predeterminadas del sistema Windows©, culminando con la activación de un segundo *script* (Figura 3). Dicho *script* desplegaba en el Escritorio de las estaciones de trabajo, mediante una ventana de comandos (CMD), una representación gráfica en forma de carabela (Figura 4), con el objetivo

de emular de manera más verosímil las consecuencias observables de un ataque real.

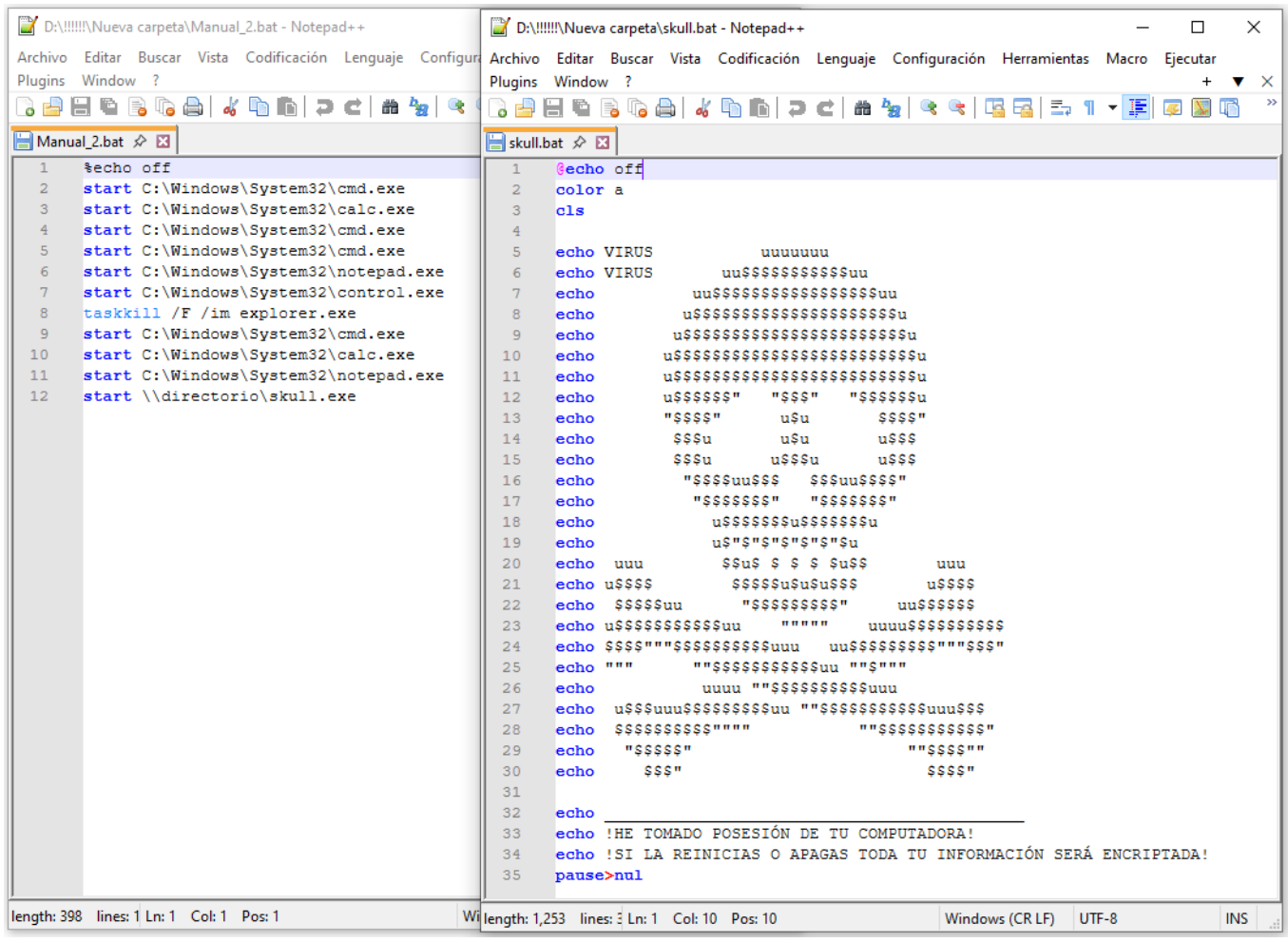

Figura 3. Primer (izquierda) y segundo (derecha) script que se ejecutaban al descargar el fichero

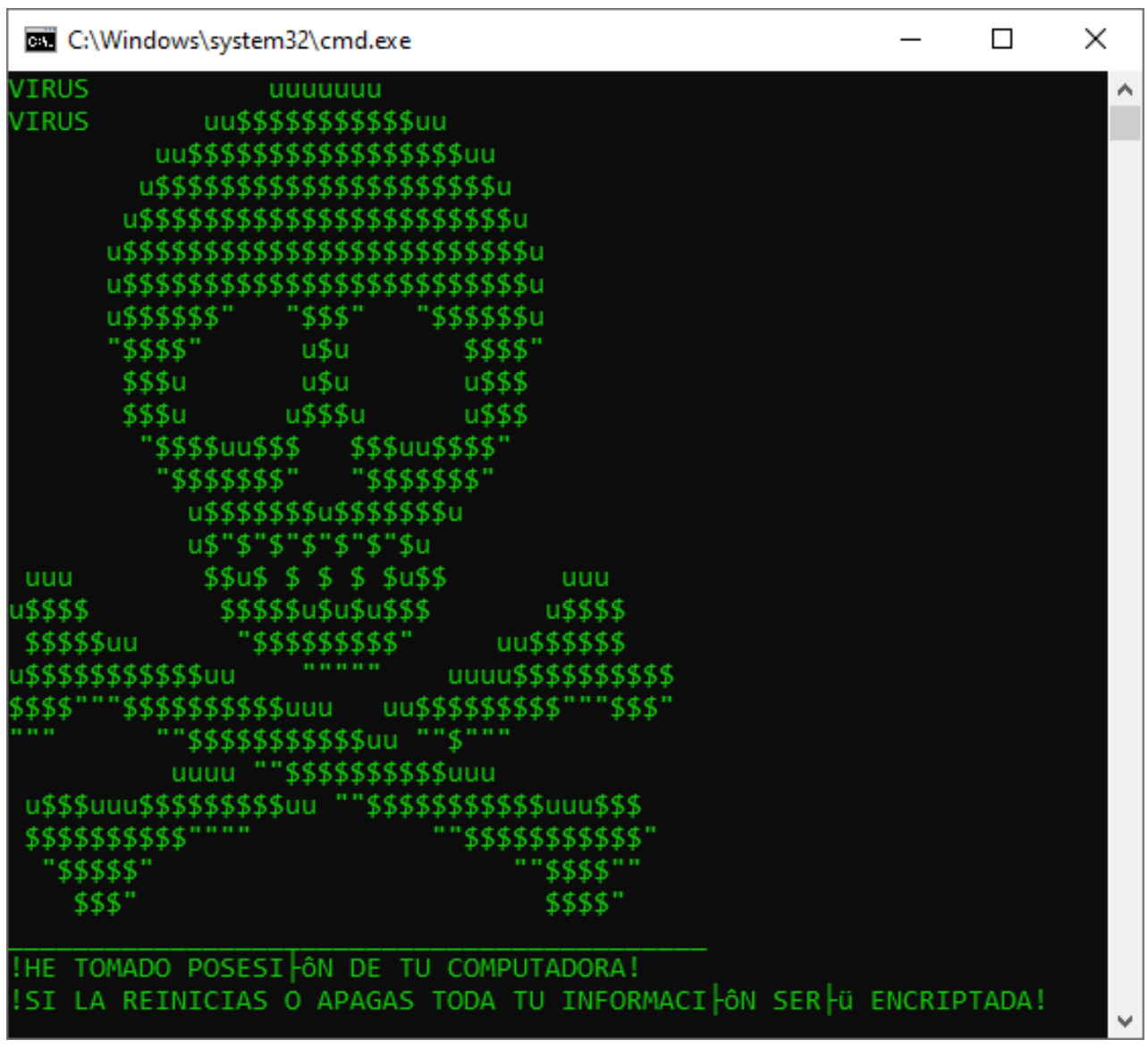

Figura 4. Representación gráfica en forma de calavera en ventana CMD

Los resultados de esta campaña fueron reveladores: si bien se registró un incremento considerable en las interacciones con el correo fraudulento (Figura 5), el número de reportes se mantuvo en ocho usuarios. Las entrevistas de seguimiento realizadas en las diferentes direcciones de la empresa identificaron la causa principal de esta

aparente discrepancia. El notable impacto visual de la simulación anterior había generado conversaciones internas entre los trabajadores, lo que resultó en un conocimiento generalizado de la prueba. En consecuencia, los usuarios que identificaron el correo optaron por no reportarlo, al reconocerlo como un elemento interno de capacitación y no como una amenaza externa genuina. Este fenómeno subraya la importancia de variar los vectores de ataque en simulaciones sucesivas para evitar la contaminación de los resultados y mantener la validez del ejercicio.

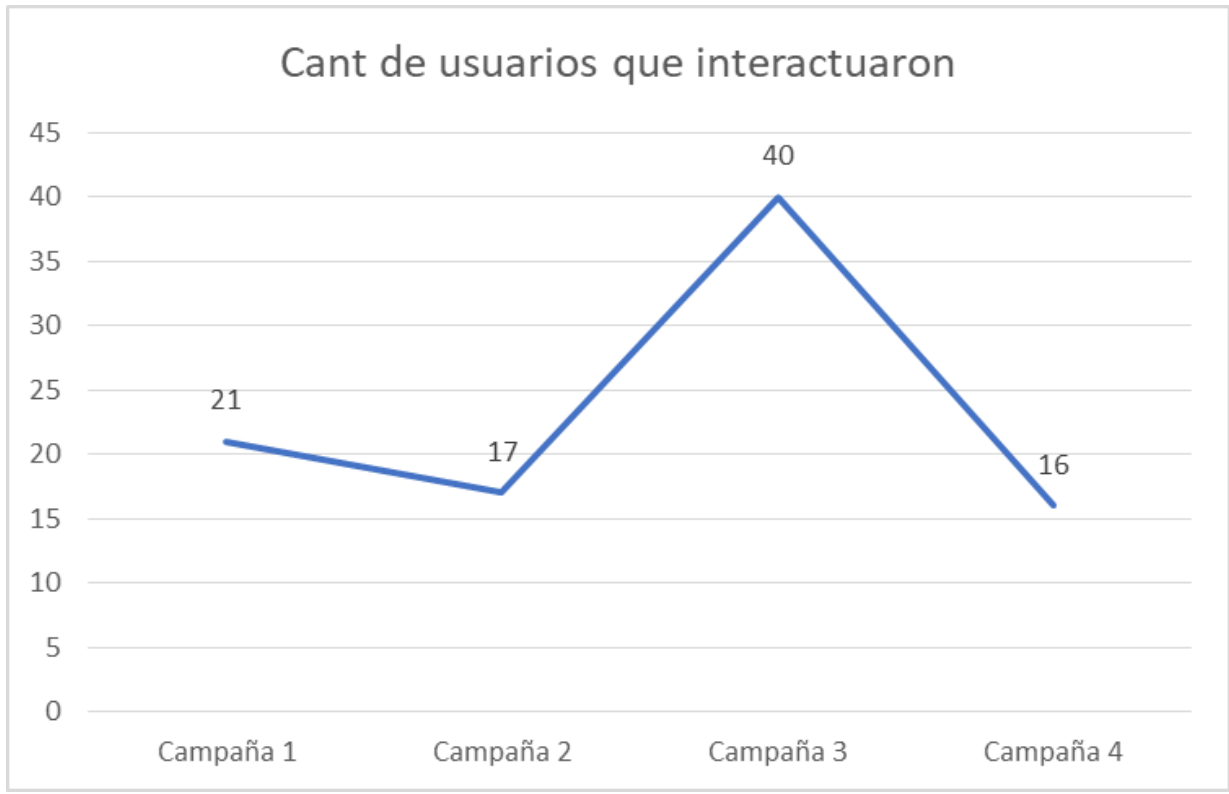

Figura 5. Cantidad de interacciones por Campañas

La Campaña 4 retomó el vector de ataque de la Campaña 1 (la solicitud de actualización de credenciales), pero aplicándolo esta vez al acceso del sistema ERP corporativo, con una plataforma de *phishing* rediseñada para esta simulación. Los resultados demostraron una mejora significativa en la resiliencia de los usuarios. Esta campaña registró el menor nivel de interacción, con 16 usuarios haciendo clic en el enlace fraudulento. De manera aún más destacable, la tasa de entrega de credenciales fue nula, ya que ninguno de los usuarios que accedió a la página falsa introdujo sus datos. Al mismo tiempo, se observó un aumento considerable en el número de reportes por parte de los empleados (Figura 6). El análisis comparativo de las cuatro campañas evidencia una clara tendencia de mejora progresiva: la reducción drástica de las interacciones y la tasa de éxito del *phishing*, junto con el incremento sostenido de las notificaciones, confirman la efectividad del programa de capacitación continuo.

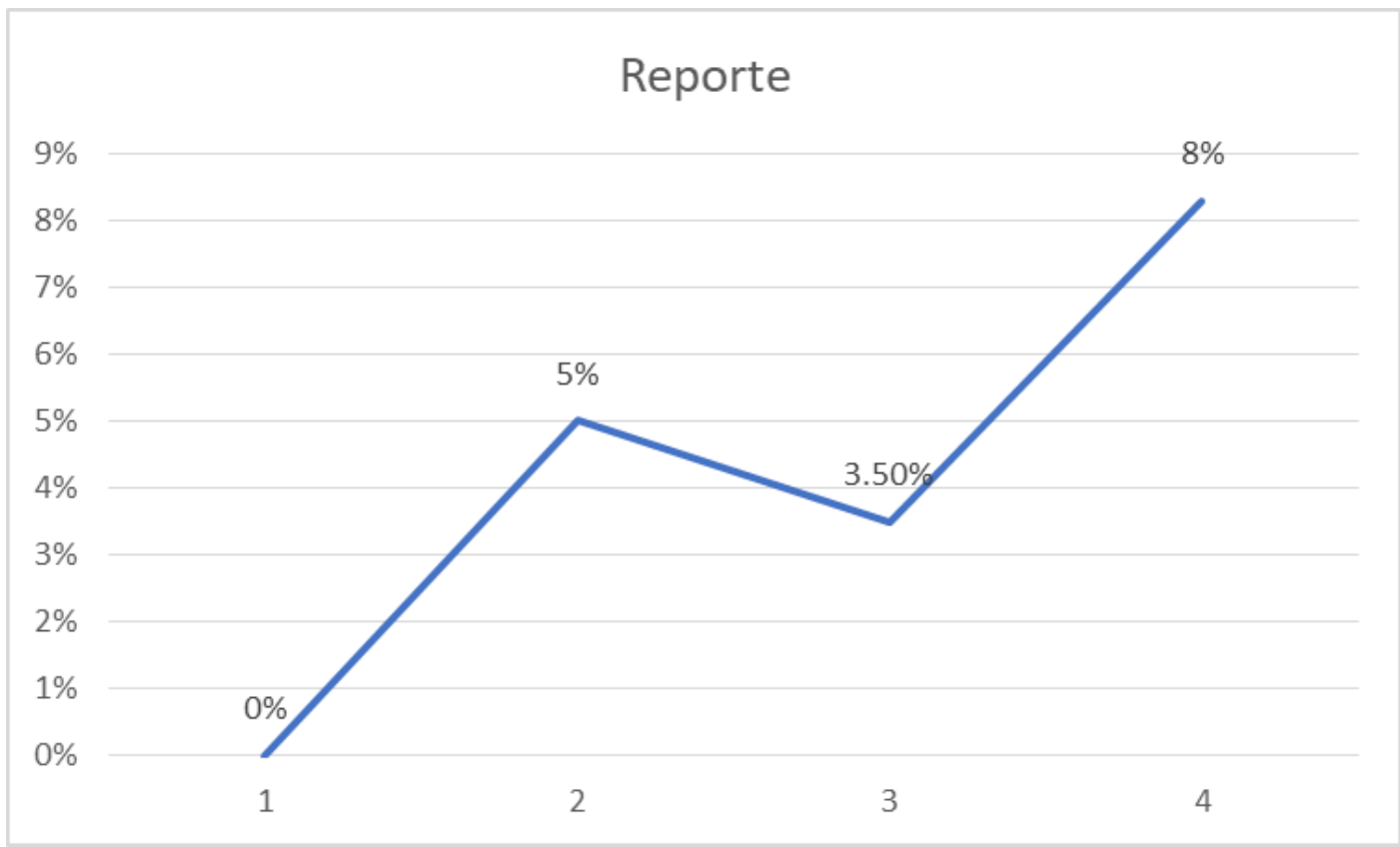

Figura 6. Reporte de recepción de los usuarios en cada Campaña

La tendencia decreciente en las tasas de interacción y entrega de credenciales a lo largo de las sucesivas campañas demuestra de manera fehaciente la efectividad del programa de capacitación implementado. Un análisis detallado de los 227 usuarios participantes en total en todas las simulaciones revela que la mayoría mostró una clara curva de aprendizaje: mientras 46 usuarios (20.3 %) solo interactuaron en una simulación, solo 21 (9.3 %) lo hicieron en dos ocasiones y solo 2 usuarios (0.9 %) mantuvieron un comportamiento vulnerable en tres de las cuatro campañas (Figura 7). Esta distribución evidencia que el entrenamiento práctico y la retroalimentación inmediata resultaron efectivos para la inmensa mayoría del personal.

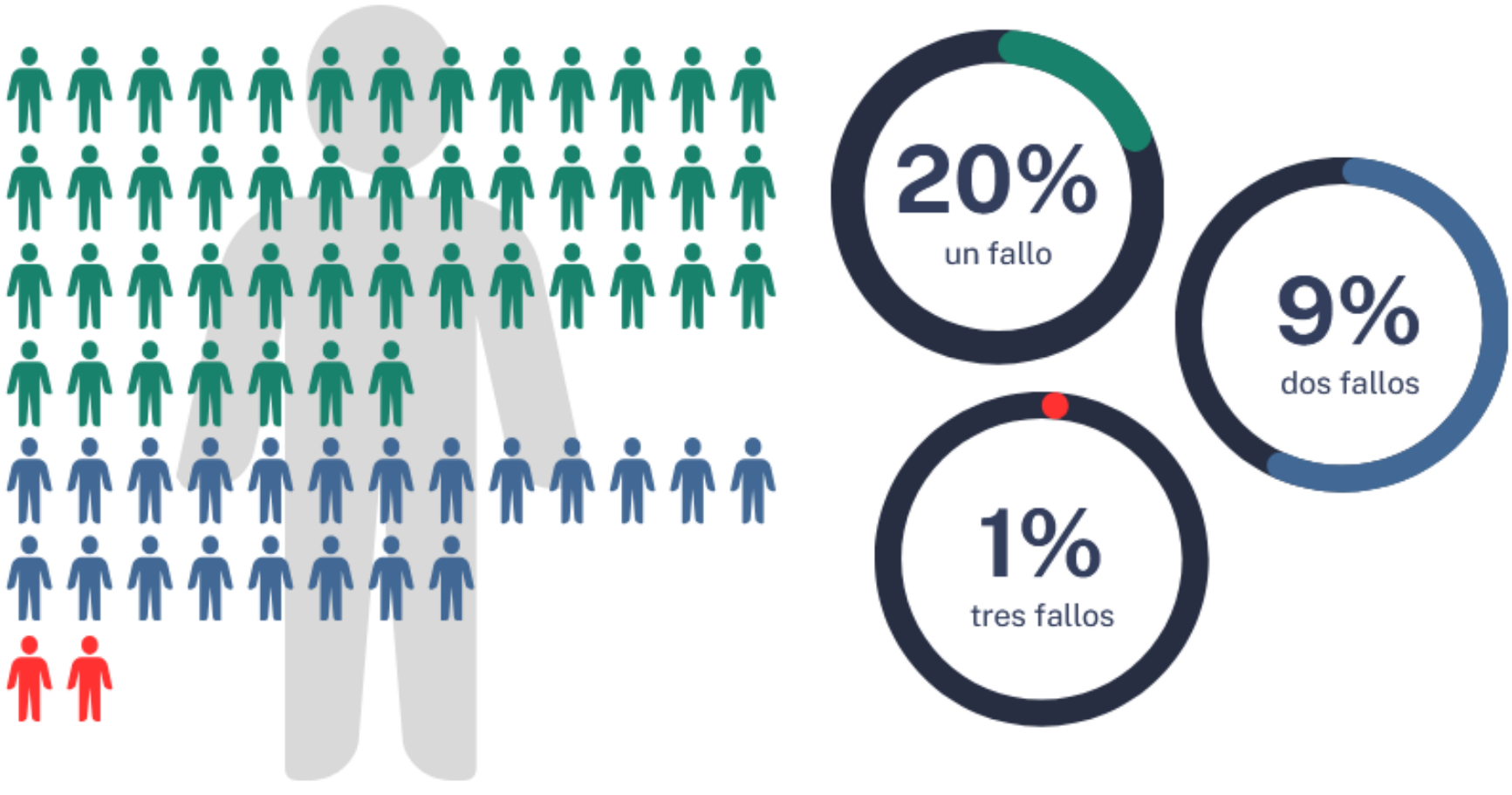

Figura 7. Cantidad de fallos reiterados por los usuarios

El éxito se ve reforzado por un indicador conductual clave: la Dirección de IT reportó un aumento sustancial en los reportes voluntarios de correos sospechosos por parte de los empleados en el entorno productivo, que trasciende el ámbito de las simulaciones. Un análisis más detallado revela que las direcciones con mayor volumen de reportes de correos *phishing* reales fueron las de Logística y Comercial, lo cual es coherente con su alta exposición a comunicaciones internacionales. Este hallazgo no solo valida la eficacia de la concientización, sino que apunta a una asimilación palpable de los protocolos de seguridad por parte de los usuarios y se traduce en un fortalecimiento de la cultura de ciberseguridad organizacional. Como resultado directo de esta mayor vigilancia, se ha incrementado la efectividad en el bloqueo proactivo de dominios maliciosos, cerrando así un ciclo de mejora continua en la postura de seguridad de la empresa.

Conclusiones

Los hallazgos de este caso de estudio, en línea con las estadísticas globales, demuestran que los enfoques de capacitación teórica tradicional resultan insuficientes para mitigar de manera eficaz el riesgo asociado al *phishing*.

La implementación de herramientas de simulación como GoPhish se revela como una estrategia crucial, pues permite cuantificar de forma objetiva el riesgo al transformar vulnerabilidades abstractas en métricas tangibles. Esta capacidad de presentar datos concretos (como el porcentaje específico de empleados que interactúan con un simulacro) facilita la comunicación efectiva con la alta dirección y la necesidad de priorizar recursos de seguridad.

Además, estas herramientas posibilitan una capacitación altamente dirigida y una optimización de recursos, al identificar con precisión a los usuarios más vulnerables. Este enfoque permite diseñar programas proactivos de concientización personalizados que maximiza el impacto de las iniciativas de entrenamiento. De manera paralela, las simulaciones continuas fomentan un estado de alerta constante y una cultura de seguridad, lo que integra la ciberseguridad en la rutina

laboral y transforma al capital humano de un eslabón pasivo en un componente activo de la defensa organizacional.

El costo de implementar una solución como GoPhish es mínimo comparado con el costo potencial de pérdidas financieras y de reputación de una sola brecha de datos exitosa. Invertir en la preparación del factor humano es, por tanto, una de las estrategias de mitigación de riesgo más efectivas disponibles.

La fortaleza cibernética de una organización ya no puede depender únicamente de *firewall* y *software*. La defensa debe ser basada en las personas también. Las herramientas de capacitación son fundamentales para lograr la operatividad de esta visión y esto facilita a las empresas transformar a su fuerza laboral de un objetivo potencial en una barrera de defensa robusta y consciente, elevando así su postura de seguridad digital de manera significativa y sostenible.

Referencias bibliográficas

Budge, J., Blankenship, J., Sjoblom, S., & Nagel, B. (2022). The Forrester Wave™: Security Awareness and Training Solutions, Q1 2022. *Forrester Research*. Disponible en https://www.livingsecurity.com/hubfs/Forrester%20Wave%20Report%202022%20-%20Security%20Awareness%20and%20Training%20-%20Q1%202022.pdf

ENISA. (2022). ENISA Threat Landscape 2022. *European Union Agency for Cybersecurity*. Disponible en https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022

Hadlington, L. (2017). Human Factors in Cybersecurity; Examining the Link Between Internet Addiction, Impulsivity, Attitudes Towards Cybersecurity, and Risky Cybersecurity Behaviours. *Heliyon, 3 (2017)*, Artículo No-e00346. https://doi.org/10.1016/j.heliyon.2017.e00346

Hernández Domínguez, A. (2025). *Más allá de la tecnología: Apostar por las personas en la estrategia de confianza cero*. Disponible en http://www.cubadebate.cu/especiales/2025/02/21/mas-alla-de-la-tecnologia-apostar-por-las-personas-en-la-estrategia-de-confianza-cero/

Pattinson, M., Jerram, C., Parsons, K., McCormac, A., & Butavicius, M. (2015). Why do some people manage phishing e-mails better than others? *Information Management & Computer Security*, *23*(3), 270-281.

Verizon (2025). 2025 Data Breach Investigations Report. V*erizon Business*. Disponible en https://www.verizon.com/business/resources/T163/reports/2025-dbir-data-breach-investigations-report.pdf

Workman, M. (2008). Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security. *Journal of the American Society for Information Science and Technology, Volumen* 59, Issue 4, pages 662-674. Disponible en https://sci-hub.se/10.1002/asi.20779

INVESTIGACIÓN

Herramienta para el monitoreo y análisis de trazas en la Fiscalía Provincial de Mayabeque

Tool for monitoring and trace analysis in the Provincial Prosecutor's Office of Mayabeque

TS. Laura Celín Sosa Martínez*

Recibido: 11/2025 | Aceptado: 11/2025 | Publicado: 12/2025

Resumen

Este estudio abordó la implementación de una herramienta especializada para el monitoreo y análisis de trazas, con el objetivo de fortalecer la supervisión de los sistemas informáticos y reducir la exposición ante amenazas. Se establecieron criterios para la selección e integración de la herramienta, priorizando la centralización de registros, el análisis en tiempo real y la visualización accesible de la información. La solución fue implementada en un entorno controlado, lo que permitió la detección de eventos relevantes, la identificación de patrones anómalos y la clasificación de amenazas según su nivel de impacto. La validación de la herramienta evidenció mejoras significativas en la capacidad para interpretar datos críticos y responder de forma oportuna ante posibles fallos, optimizando así la supervisión de la infraestructura tecnológica. Los resultados obtenidos confirmaron que el análisis de trazas con soporte visual representa una estrategia efectiva para garantizar la continuidad operativa y la protección de los sistemas en instituciones públicas.

Palabras clave: amenazas, análisis de trazas, monitoreo, visualización de datos, seguridad informática

* Dirección de Seguridad Informática de la UCI, Carretera a San Antonio de los Baños, Km 2 ½, reparto Torrens, municipio Boyeros, La Habana, Cuba.lsosa@uci.cu.

Abstract

This study addressed the implementation of a specialized tool for monitoring and analyzing traces, with the goal of strengthening the oversight of information systems and reducing exposure to threats. Criteria were established for the selection and integration of the tool, prioritizing the centralization of logs, real-time analysis, and accessible visualization of information. The solution was implemented in a controlled environment, allowing for the detection of relevant events, identification of anomalous patterns, and classification of threats based on their level of impact. The validation of the tool demonstrated significant improvements in the ability to interpret critical data and respond promptly to potential failures, thus optimizing the supervision of technological infrastructure. The results confirmed that trace analysis with visual support represents an effective strategy for ensuring operational continuity and protecting systems in public institutions.

Keywords: *threats, trace analysis, monitoring, data visualization, cybersecurity*

Introducción

El monitoreo y análisis de trazas se ha vuelto una práctica esencial para el correcto funcionamiento de los sistemas informáticos en instituciones públicas y privadas. En el contexto actual, donde los sistemas informáticos están cada vez más interconectados y dependen de la tecnología para operar, es fundamental poder supervisar las actividades de las redes y los servidores para identificar problemas de rendimiento, fallos en los procesos y posibles amenazas (Alzahrani et al., 2025).

A medida que las organizaciones se digitalizan y dependen más de los sistemas tecnológicos; la seguridad de la información y la eficiencia operativa se convierten en prioridades claves. Las trazas registran información detallada sobre cada evento dentro de los sistemas, lo que permite no solo realizar diagnósticos precisos de los problemas, sino también asegurar la confidencialidad e integridad de los datos (Ahmed et al., 2023).

En el caso de Cuba, el monitoreo y análisis de trazas se ha vuelto una prioridad creciente a medida que el país avanza en su proceso de informatización de la sociedad. Con la adopción de nuevas tecnologías y la digitalización de los servicios públicos, la necesidad de implementar soluciones efectivas para monitorear los sistemas informáticos es más importante que nunca. A pesar de los desafíos tecnológicos y de conectividad, Cuba ha logrado avances significativos en el fortalecimiento de su infraestructura tecnológica, priorizando la seguridad y el desempeño de sus sistemas (Álvarez & Andrés, 2023).

La Fiscalía Provincial de Mayabeque, fundada en el año 2010, tiene como objetivo velar por el cumplimiento de la Constitución, las leyes y demás disposiciones legales, tanto de los organismos del Estado como de los ciudadanos. Su misión es ejercer el control de la investigación penal y la acción penal pública en representación del Estado, asegurando el estricto cumplimiento normativo, protegiendo los datos y la información. Esto implica un compromiso con la seguridad y la privacidad de la información manejada, todo lo anterior garantiza no solo la transparencia en sus funciones, sino también el respeto a los derechos de las personas.

En la Fiscalía Provincial de Mayabeque se evidenció una grave deficiencia en el monitoreo de trazas. Uno de los hallazgos más preocupantes fue la falta de un sistema efectivo para rastrear y documentar las trazas de información, lo que afecta directamente la capacidad de la Fiscalía para detectar anomalías y fallos recurrentes en sus sistemas informáticos, esto pone en riesgo la eficiencia de los servicios que ofrece.

A partir de lo antes expuesto se plantea el siguiente problema de investigación: ¿Cómo contribuir al mejoramiento del monitoreo y análisis de trazas en la Fiscalía Provincial de Mayabeque?

Para dar solución a dicho problema, se tiene como objetivo general: implementar una herramienta para el monitoreo y análisis de trazas en la Fiscalía Provincial de Mayabeque con el fin de detectar anomalías.

Materiales y métodos

Para seleccionar la herramienta adecuada para el monitoreo y análisis de trazas, se realizó una búsqueda exhaustiva de soluciones tecnológicas reconocidas por su capacidad de análisis avanzado y su eficacia en la detección de problemas. Entre las opciones evaluadas se identificaron Grafana, Prometheus y ELK Stack, todas de código abierto y ampliamente utilizadas en entornos de observabilidad y supervisión de sistemas.

Grafana es una herramienta de visualización que analiza datos en tiempo real mediante paneles interactivos y personalizables. Se integra con múltiples fuentes como Prometheus, Elasticsearch e InfluxDB y ofrece funcionalidades como alertas configurables, integración con OpenTelemetry y diseño de paneles gráficos (Leppänen, 2021).

Prometheus, por su parte, está orientada a la recopilación de métricas en forma de series temporales. Su modelo de datos multidimensional y su lenguaje de consulta PromQL permite realizar análisis detallados y generar alertas personalizadas. Aunque su visualización es limitada, se complementa de manera eficaz con Grafana (Kim et al., 2024).

ELK Stack se destaca por su capacidad para centralizar y analizar trazas de múltiples fuentes en sistemas distribuidos. Su arquitectura escalable gestiona grandes volúmenes de datos, correlaciona eventos en tiempo real y detecta patrones de comportamiento. Además, facilita la visualización detallada mediante Kibana, lo que mejora la interpretación de datos y la toma de decisiones (Gatsi et al., 2023).

Para responder a las necesidades específicas de la Fiscalía Provincial de Mayabeque, se definieron criterios clave para la selección de la herramienta: compatibilidad con los sistemas existentes, costo nulo por ser de código abierto, capacidad avanzada de análisis de trazas, eficiencia en la búsqueda de datos, facilidad de uso para personal no técnico y amplias opciones de visualización.

A continuación, se presenta una tabla comparativa (Tabla 1) que resume las características de las herramientas evaluadas:

Tabla 1. Comparación de las herramientas para monitoreo de trazas

Herramientas	Compatibilidad	Costo	Análisis de Trazas	Búsqueda de Datos	Usabilidad	Visualización
Grafana	Se integra con Windows, Linux y MacOS.	Gratuito	Limitado, depende de integración externa.	No realiza búsquedas directas.	Interfaz gráfica personalizable.	Interfaz gráfica personalizable.
Prometheus	Se integra con Windows, Linux y MacOS.	Gratuito	Buena capacidad para patrones complejos.	Consultas avanzadas.	Requiere conocimientos técnicos.	Visualización limitada, se complementa con Grafana.
ELK Stack	Se integra con Windows, Linux y MacOS.	Gratuito	Alta capacidad para detectar patrones.	Búsquedas rápidas y eficientes.	Interfaz intuitiva y flexible.	Visualización dinámica y en tiempo real.

Con base en esta comparación, se determinó que ELK Stack es la herramienta más adecuada para implementar en la Fiscalía Provincial de Mayabeque. Su versatilidad, capacidad de análisis, eficiencia en la búsqueda de datos y facilidad de uso la convierten en la solución ideal para mejorar el monitoreo de trazas.

La interfaz gráfica de Kibana permite una adopción rápida por parte del personal, mientras que la integración con Logstash y Elasticsearch garantiza una centralización efectiva de los registros. ELK Stack cumple con todos los requisitos establecidos, y ofrece una plataforma robusta para la supervisión de sistemas, la detección de anomalías y la generación de informes visuales que facilitan la toma de decisiones informadas.

Resultados y discusión

La herramienta ELK Stack fue implementada en un entorno controlado dentro de la Fiscalía Provincial de Mayabeque, con el propósito de validar su efectividad en el monitoreo y análisis de trazas. Se configuraron los componentes principales: Logstash, para la recolección y procesamiento de datos; Elasticsearch, para el almacenamiento y búsqueda eficiente; y Kibana, para la visualización interactiva de los registros.

Durante la implementación, se definieron variables de entorno específicas (Figura 1) que permitieron adaptar la herramienta a las características de la infraestructura tecnológica de la Fiscalía. Se verificó el funcionamiento de los índices generados por Elasticsearch

y se activó el servidor de Kibana para la visualización en tiempo real de los eventos registrados.

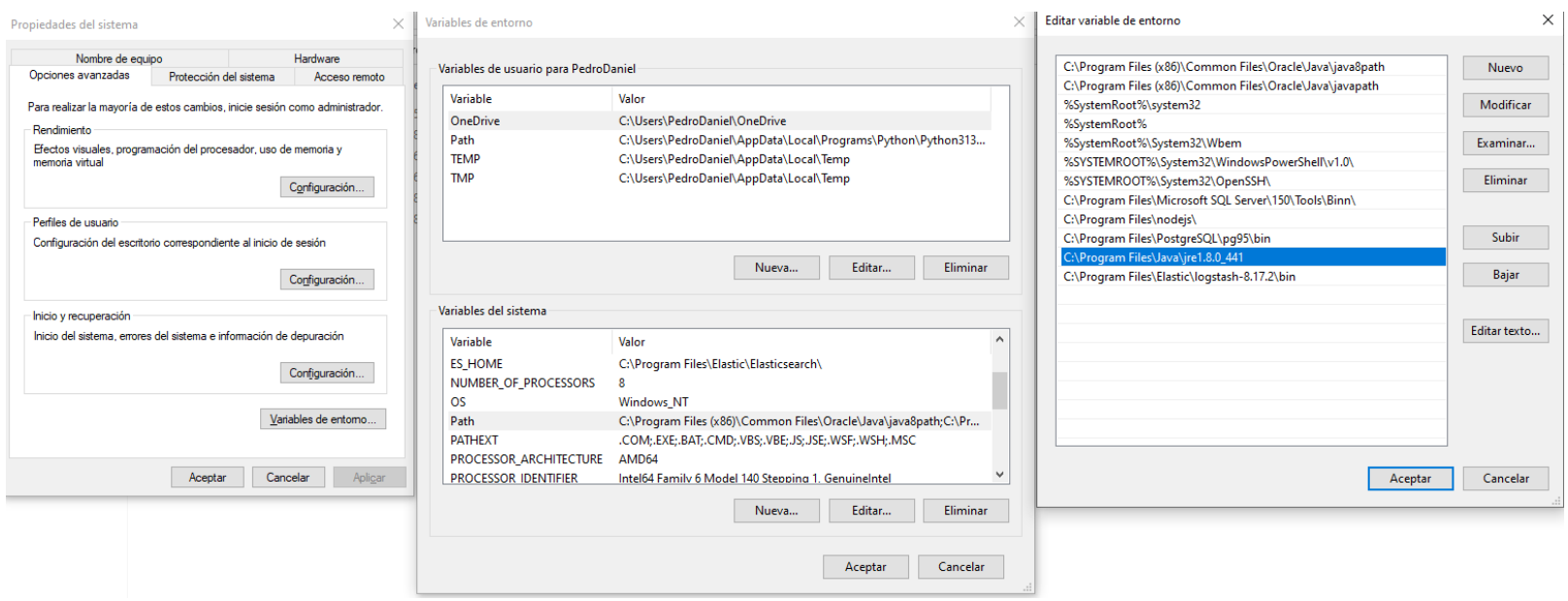

Figura 1. Variables de entorno

Una vez completada la configuración, se procedió a la validación funcional de la herramienta. Se realizaron pruebas de detección de eventos, comprobación de alertas y análisis de trazas en distintos escenarios simulados (Figura 2). Los resultados mostraron que la herramienta fue capaz de identificar patrones anómalos, clasificar amenazas según su nivel de impacto y generar alertas oportunas ante posibles fallos (Figura 3 y Figura 4).

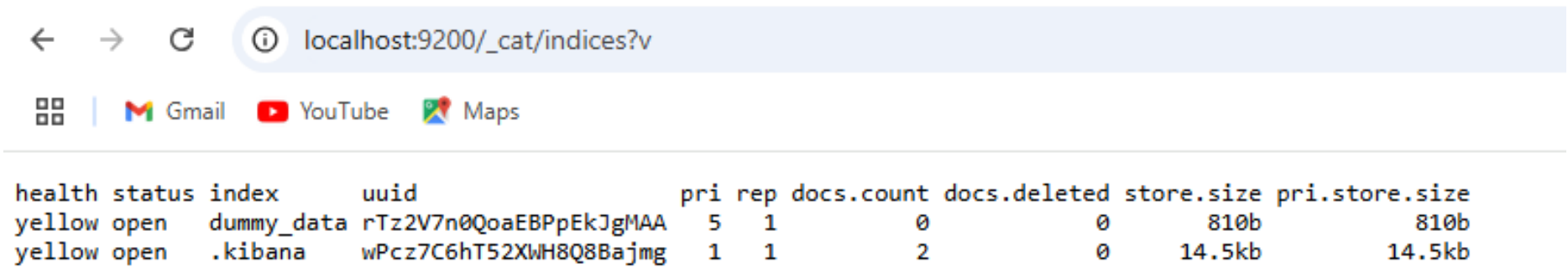

Figura 2. Comprobación de los índices

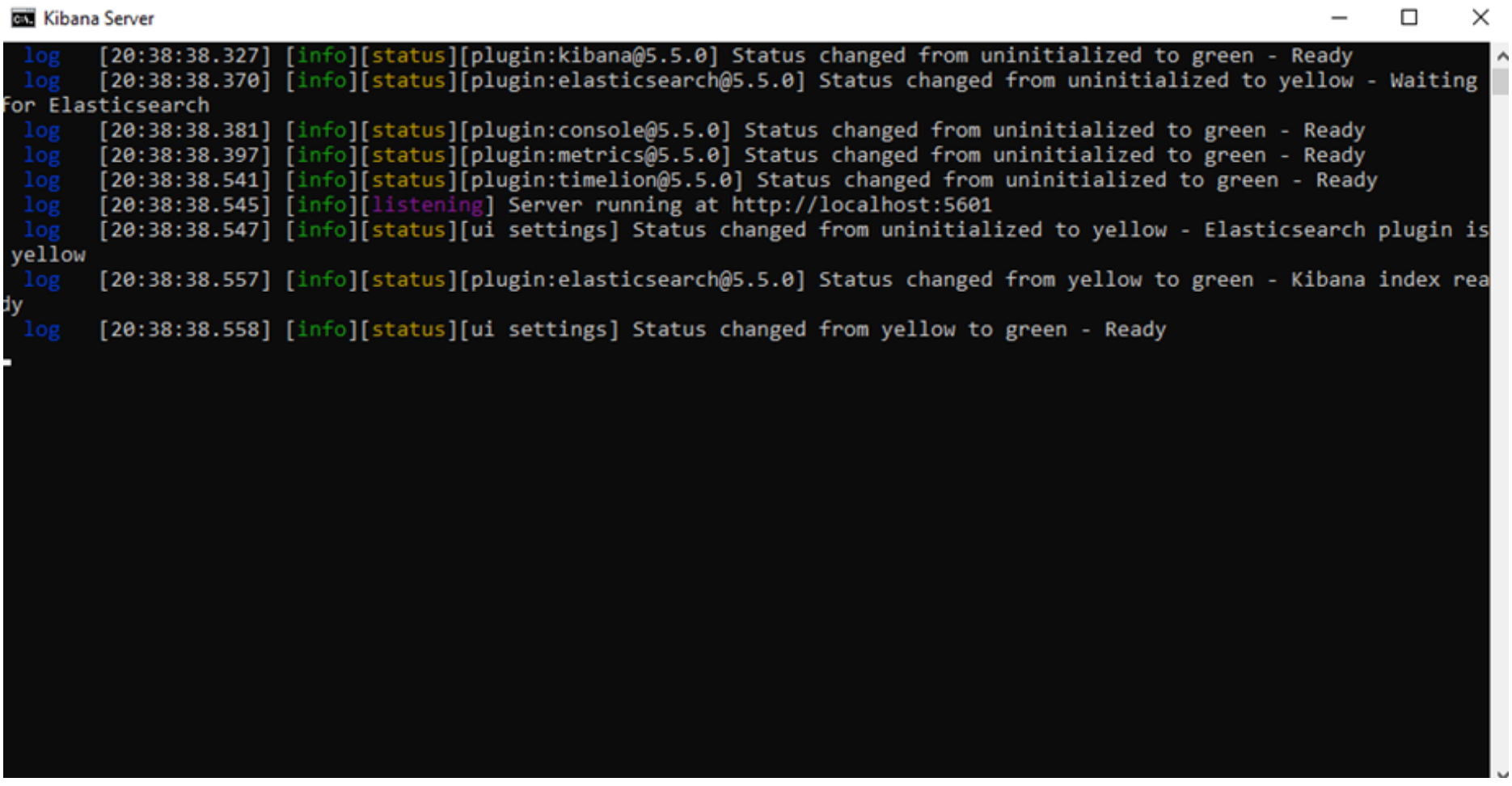

Figura 3. Servidor de Kibana activo

logstash.conf.txt: Bloc de notas

Archivo Edición Formato Ver Ayuda

```
input {
     file {
          path => "Escritorio/dummy_data.log"
          start_position => "beginning"

     }
}

filter {
     json {
     source => "message"
   }
}

output {
     elasticsearch {
          hosts => ["http://localhost:9200"]
          index => "dummy_data"
     }
stdout {codec => rubydebug }
}
```

Figura 4. Configuración del Logstash

La visualización de los eventos detectados permitió al personal técnico interpretar los datos de forma clara y rápida, lo anterior facilitó la toma de decisiones informadas. Además, se logró una mejora significativa en la capacidad de respuesta ante incidentes y esto contribuyó a optimizar la supervisión de los sistemas informáticos (Figura 5; Figura 6; Tabla 2; Tabla 3; Tabla 4).

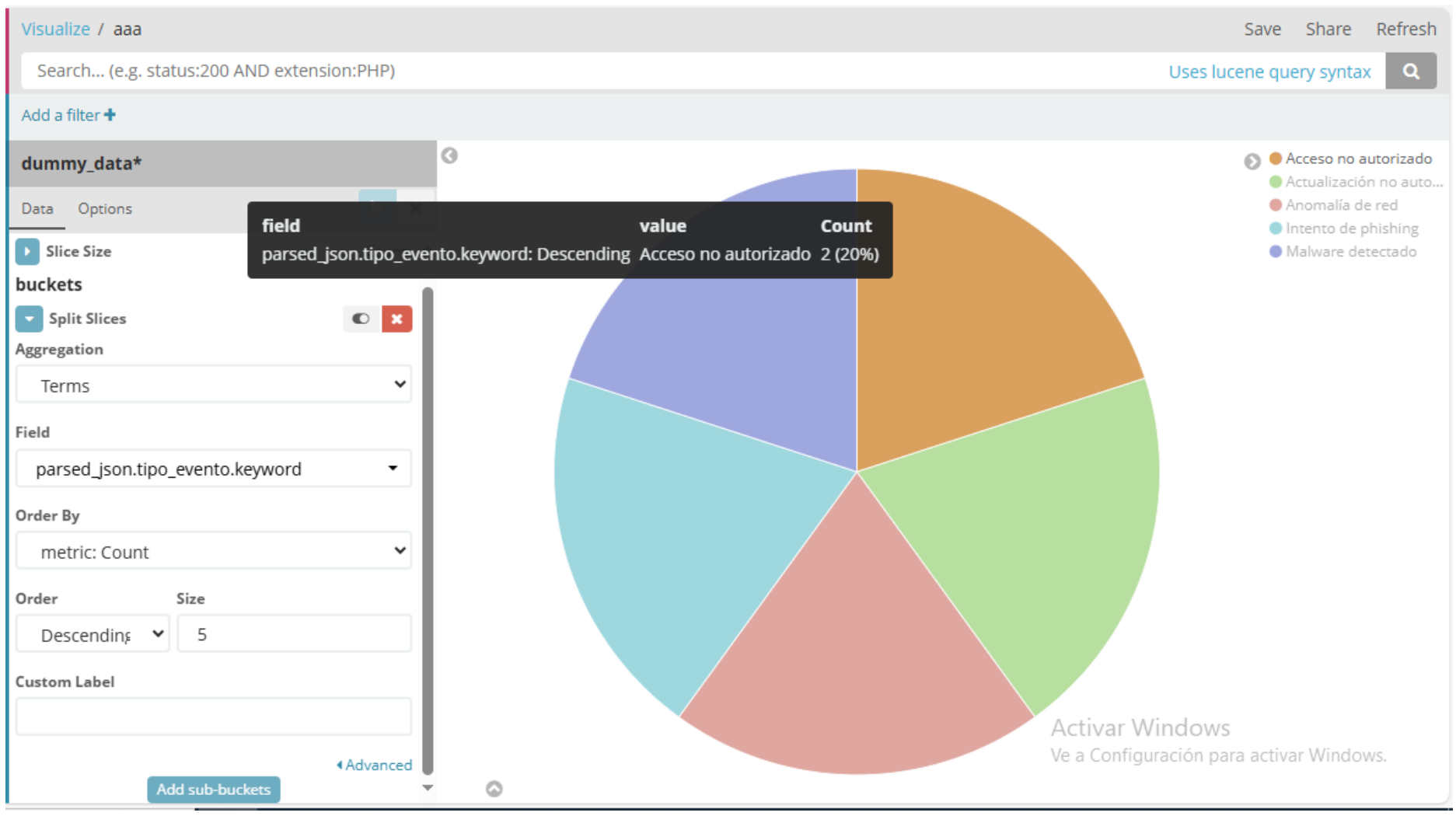

Figura 5. Eventos detectados

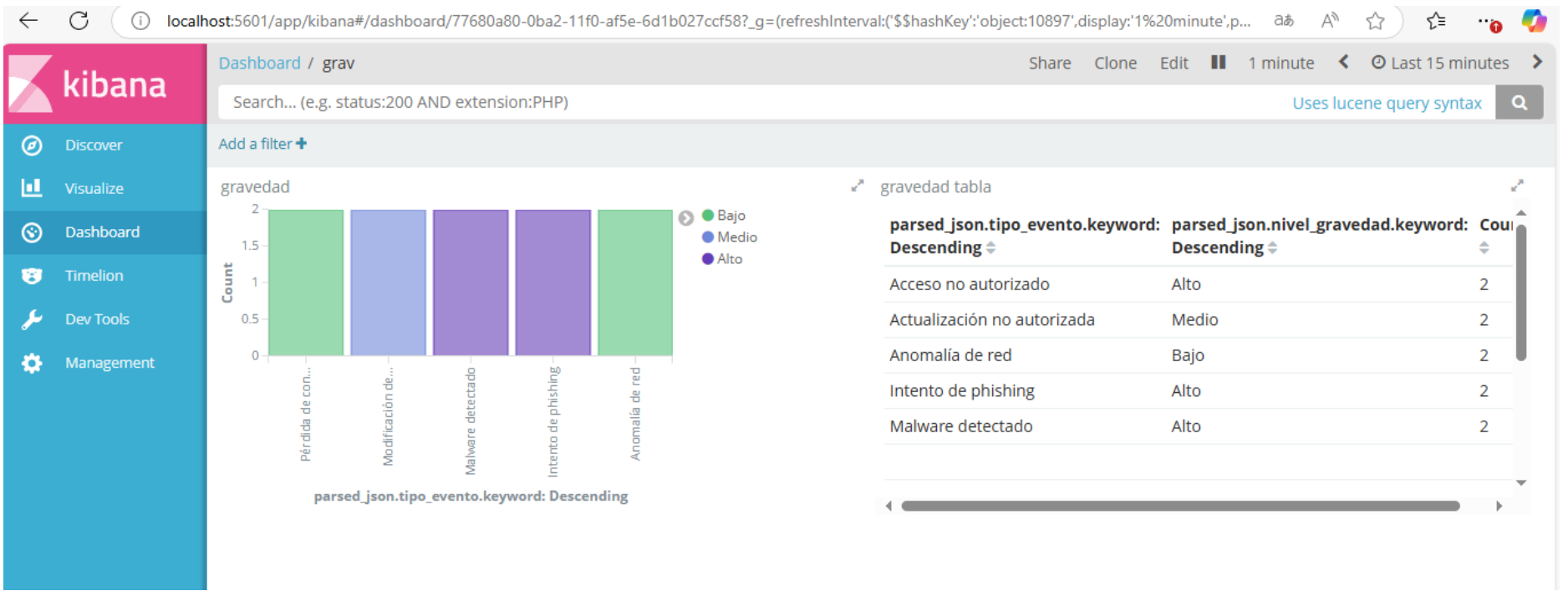

Figura 6. Clasificación de amenazas

Tabla 2. Clasificación de eventos detectados

Evento	Cantidad Detectada	Gravedad	Acción Realizada
Intentos de acceso no autorizado.	15	Crítico	Bloqueo de IP y ajustes de *firewall*.
Descargas sospechosas.	10	Medio	Restricción de descargas.
Actividades sospechosas.	5	Bajo	Monitoreo continuo.

Tabla 3. Clasificación de amenazas identificadas

Amenaza Identificada	Descripción	Clasificación de Gravedad	Acciones Recomendadas
Servicio expuesto vulnerable	Configuración deficiente que permite acceso no autorizado desde Internet.	Crítico	Actualizar configuraciones y aplicar políticas estrictas
Descargas no autorizadas	Descargas que pueden introducir *malware* al sistema sin intervención de usuarios.	Medio	Restringir descargas.
Ataques basados en ingeniería social	Aplicaciones maliciosas entregadas por correo electrónico o descargadas por usuarios.	Bajo	Implementar filtros de correo y educar a usuarios.

Tabla 4. Resumen de problemas y resoluciones

Problema	Acción Correctiva	Estado
Servicio vulnerable expuesto.	Reconfiguración y aplicación de restricciones.	Resuelto
Actividades sospechosas.	Capacitación de usuarios y monitoreo continuo.	En proceso

Los resultados obtenidos evidencian que la solución propuesta constituye una mejora sustancial respecto a las prácticas anteriores. La centralización de registros, el análisis avanzado de trazas y la visualización dinámica representan una novedad en el contexto de la Fiscalía Provincial de Mayabeque, donde no existía una herramienta con estas capacidades.

ELK Stack demostró ser superior a otras soluciones similares en términos de funcionalidad, adaptabilidad y facilidad de uso, lo que confirma el cumplimiento de los objetivos planteados en este estudio. Su implementación no solo fortaleció la seguridad informática de la institución, sino que también sentó las bases para futuras mejoras en la gestión tecnológica.

Conclusiones

La implementación de la herramienta ELK Stack en la Fiscalía Provincial de Mayabeque mejoró de forma significativa el monitoreo y análisis de trazas en sus sistemas informáticos. Esta solución tecnológica respondió de manera efectiva al problema identificado en la auditoría, al proporcionar una plataforma capaz de centralizar registros, detectar eventos relevantes y visualizar datos de forma clara y dinámica.

Se cumplió el objetivo general del estudio, al establecer una herramienta que facilita la identificación de anomalías, fortalece la seguridad de la infraestructura tecnológica y optimiza la supervisión operativa. Los resultados obtenidos evidencian que el uso de trazas con soporte visual representa una estrategia eficiente para garantizar la continuidad de los servicios institucionales.

El trabajo realizado constituye un aporte relevante en el campo de la administración de redes y seguridad informática, al demostrar que es posible implementar soluciones avanzadas con recursos accesibles y de código abierto. La experiencia adquirida en este proceso puede servir de referencia para otras instituciones públicas que enfrenten desafíos similares en la gestión de sus sistemas tecnológicos.

Como proyección futura, se recomienda continuar el desarrollo de la herramienta incorporando mecanismos de inteligencia artificial para la detección predictiva de amenazas, así como ampliar su integración con otros sistemas de monitoreo existentes. Además, se sugiere capacitar al personal técnico en el uso de estas tecnologías para garantizar su aprovechamiento pleno y sostenible.

Referencias bibliográficas

Ahmed, A., Khan, M., & Hussain, S. (2023). Trace analysis for anomaly detection in distributed systems. *Journal of Computer Networks and Security*, 15(2), 45–58.

Álvarez, J., & Andrés, R. (2023). Seguridad informática en entornos gubernamentales: desafíos y soluciones. *Revista Cubana de Tecnología*, 29(1), 12–20.

Gatsi, T., Mensah, K., & Boateng, E. (2023). Enhancing system observability using ELK Stack in large-scale infrastructures. *International Journal of Information Systems*, 18(4), 101–115.

Haykal, A., & Siswanto, R. (2020). Monitoring and threat detection in cloud-native environments. *Cybersecurity Review*, 7(3), 88–97.

Kim, J., Park, S., & Lee, H. (2024). Prometheus-based monitoring for scalable cloud applications. *Journal of Cloud Computing*, 22(1), 33–49.

Leppänen, M. (2021). Grafana dashboards for real-time system analytics. *Software Engineering Insights*, 10(2), 67–74.

INVESTIGACIÓN

Aplicación de modelos de inteligencia artificial para la detección de malware en infraestructuras críticas de telecomunicaciones

Implementing artificial intelligence models for malware detection in critical telecommunications infrastructures

Ing. Elizabeth Molina Mena[*], Ing. Heidy Rodríguez Malvares[2], M.Sc. Henry Raúl González Brito[3]

Recibido: 11/2025 | Aceptado: 11/2025 | Publicado: 12/2025

Resumen

El avance acelerado de la digitalización y la interconexión global ha incrementado la exposición de las infraestructuras críticas de telecomunicaciones a diversas amenazas cibernéticas, entre las cuales el *malware* representa una de las más persistentes y dañinas. En este contexto, la aplicación de modelos de inteligencia artificial (IA) para la detección temprana de *software* malicioso se ha convertido en una estrategia clave para garantizar la resistencia, la disponibilidad y la continuidad de los servicios esenciales. Este artículo analiza los principales enfoques basados en aprendizaje automático y aprendizaje profundo empleados en la identificación de *malware*,

[*] Dirección de Seguridad Informática, Universidad de las Ciencias Informáticas. elizabethmm@uci.cu

[2] Departamento de Infraestructuras Tecnológicas, Facultad de Ciberseguridad, Universidad de las Ciencias Informáticas. heidyrm@uci.cu

[3] Dirección de Seguridad Informática, Universidad de las Ciencias Informáticas. henryraul@uci.cu

con énfasis en su implementación dentro de entornos de telecomunicaciones críticos.

Se revisan técnicas clásicas como Support Vector Machines (SVM) y Random Forest, así como arquitecturas avanzadas de redes neuronales, incluyendo Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) y Graph Neural Networks (GNN). La metodología se centra en una revisión sistemática de la literatura reciente y un análisis comparativo de modelos según métricas de rendimiento, capacidad de generalización y consumo de recursos computacionales. Los resultados muestran que los modelos híbridos basados en aprendizaje profundo y técnicas de análisis de flujo de red logran tasas de detección superiores al 98 % con una reducción significativa de falsos positivos. Por último, se discuten los desafíos actuales y las perspectivas futuras para la integración de IA en la ciberdefensa de infraestructuras críticas, y se destaca la importancia de la explicabilidad, la detección de amenazas desconocidas y la colaboración entre operadores e instituciones de investigación.

Palabras clave: inteligencia artificial, detección de *malware*, infraestructuras críticas, telecomunicaciones, ciberseguridad

Abstract

The rapid advance of digitization and global interconnection has made critical telecommunications infrastructure more vulnerable to various cyber threats, among which, malware is one of the most persistent and damaging. In this context, implementing Artificial Intelligence (AI) models to detect malicious software early on has become a key strategy for ensuring the resilience, availability, and continuity of essential services. This article analyzes the primary machine learning and deep learning approaches used for malware identification, focusing on their implementation in critical telecommunications environments. Classical techniques such as Support Vector Machines (SVM) and Random Forest are reviewed, as well as advanced neural network architectures, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Graph Neural Networks (GNN). The methodology focuses on a systematic review of recent literature and a comparative analysis

of models based on performance metrics, generalization capability, and computational resource consumption. Results indicate that hybrid models combing deep learning and network flow analysis techniques can achieve 98% rates detection with a significantly fewer false positive. Finally, current challenges and future perspectives for the integration of AI into the cyber defense of critical infrastructures are discussed, highlighting the importance of explainability, unknown threat detection, and collaboration between Telcos carriers and research institutions.

Keywords: *Artificial Intelligence, malware detection, critical infrastructures, telecommunications, cybersecurity*

Introducción

Las infraestructuras críticas de telecomunicaciones constituyen el núcleo de los sistemas de comunicación modernos, que facilitan los servicios esenciales para el funcionamiento de gobiernos, empresas y ciudadanos. Estas infraestructuras abarcan redes de transmisión, centros de datos, sistemas de conmutación y plataformas de gestión que garantizan la conectividad nacional e internacional. Sin embargo, su creciente complejidad y dependencia de *software* las convierten en un objetivo prioritario para los ciberatacantes (García-Teodoro et al., 2023).

El *malware* (término que engloba a virus, troyanos, *ransomware*, *spyware* y *rootkits*) ha evolucionado en sofisticación, empleando técnicas de evasión como ofuscación de código, polimorfismo y *living-off-the-land* (LOTL), esto dificulta su detección por mecanismos tradicionales basados en firmas (Alasmary et al., 2022). En respuesta a esta problemática, la inteligencia artificial (IA) se ha posicionado como una herramienta fundamental para la identificación automatizada de patrones anómalos y la detección de amenazas emergentes en tiempo real (Sgandurra & Lupu, 2020).

El aprendizaje automático (*Machine Learning*, ML) y el aprendizaje profundo (*Deep Learning*, DL) han demostrado un potencial notable en la clasificación de *malware* a partir de grandes volúmenes de datos, extrayendo características tanto estáticas (del código binario) como dinámicas (del comportamiento en ejecución). Estos modelos

aprenden representaciones complejas que identifican variantes nuevas o desconocidas, incluso en escenarios de ataques dirigidos contra infraestructuras críticas (Ucci, Aniello & Baldoni, 2019).

El presente artículo tiene como objetivo analizar la aplicación de modelos de inteligencia artificial en la detección de *malware* dentro de sectores estratégicos de telecomunicaciones. Se aborda el problema desde una perspectiva técnico-científica, y combina el estudio de algoritmos, técnicas de extracción de características, y el contexto operativo de las redes de telecomunicaciones. Además, se examinan los desafíos asociados a la interpretabilidad de los modelos, la gestión de grandes volúmenes de datos y la integración con sistemas de seguridad existentes como los *Security Information and Event Management* (SIEM) y los *Intrusion Detection Systems* (IDS).

Este trabajo se estructura de la siguiente manera: en la sección de Materiales y Métodos se describe la metodología empleada para la revisión sistemática y la clasificación de modelos de IA; en Resultados y Discusión se analizan los hallazgos principales y su aplicabilidad en entornos reales; y, en Conclusiones se resumen los aportes expuestos y se plantean futuras líneas de investigación orientadas al fortalecimiento de la ciberresiliencia en el sector de las telecomunicaciones.

Materiales y métodos

Para la elaboración de este estudio sobre la aplicación de modelos de inteligencia artificial en la detección de *malware* en infraestructuras críticas de telecomunicaciones, se adoptó una metodología de revisión sistemática y análisis comparativo de enfoques existentes, complementada con una caracterización de los modelos más representativos implementados en entornos reales de telecomunicaciones.

Fuentes de información y criterios de selección

La búsqueda bibliográfica se realizó en bases de datos científicas de alto impacto como IEEE Xplore, SpringerLink, ScienceDirect, ACM Digital Library y Scopus, utilizando combinaciones de palabras clave como «AI-based malware detection», «deep learning for cyber defense», «critical infrastructure protection», y «telecommu-

nication cybersecurity».

Se establecieron los siguientes criterios de inclusión:

- Publicaciones revisadas por pares entre 2018 y 2025.
- Estudios enfocados en detección o clasificación de *malware* mediante IA.
- Investigaciones aplicadas a infraestructuras críticas o redes de telecomunicaciones.
- Trabajos que reportaran métricas de rendimiento cuantitativas (precisión, *recall*, F1-score o tasa de falsos positivos).

Asimismo, se excluyeron estudios con conjuntos de datos no verificables o con modelos sin reproducibilidad documentada. En total, se revisaron 96 publicaciones científicas, de las cuales 38 cumplieron los criterios establecidos y fueron incluidas en el análisis principal.

Metodología de análisis

La metodología adoptada se estructuró en tres fases:

1. Revisión sistemática de literatura: se empleó el protocolo PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) para identificar y filtrar los estudios relevantes. La búsqueda inicial generó 327 resultados, que se redujeron tras la aplicación de los filtros de calidad, año y ámbito de aplicación.

2. Clasificación de modelos de IA: los modelos se agruparon según su naturaleza algorítmica en cuatro categorías principales:
   - Modelos clásicos de ML: SVM, Decision Trees, Random Forest, Naïve Bayes, KNN.
   - Modelos de DL: CNN, RNN, LSTM, Autoencoders, GNN.
   - Modelos híbridos: combinaciones de ML y DL, como CNN+SVM o Autoencoder+Random Forest.
   - Modelos de detección basados en flujo de red y análisis de comportamiento, especialmente en entornos 5G y SDN (*Software Defined Networking*).

3. Evaluación comparativa: se analizaron las métricas de rendimiento de los modelos, comparando precisión, sensibilidad y tiempo de inferencia. Se utilizó un esquema de ponderación basado en la media geométrica de las métricas de precisión y *recall*, para determinar la eficiencia relativa de cada modelo (Zhang et al., 2022).

Infraestructuras críticas y contexto de aplicación

En el ámbito de las telecomunicaciones, las infraestructuras críticas comprenden redes troncales de fibra óptica, sistemas de señalización SS7 y SIP, nodos de conmutación, estaciones base, *routers* de *backbone* y plataformas de virtualización de funciones de red (NFV). Estas infraestructuras soportan servicios esenciales como telefonía, datos, radiocomunicación y redes satelitales (ENISA, 2023).

Debido a su carácter estratégico y dependencia intersectorial, un ataque de *malware* puede generar efectos en cascada sobre servicios energéticos, financieros y gubernamentales (CISA, 2022). Por ello, la detección temprana de anomalías mediante IA se considera un componente esencial de la ciberresiliencia operacional (Shafiq et al., 2020).

Modelos de aprendizaje automático empleados

Los algoritmos clásicos de aprendizaje supervisado como *Support Vector Machine* (SVM) y *Random Forest* (RF) han mostrado un buen rendimiento en la detección de *malware* mediante la extracción de características estáticas (Permana et al., 2020).

SVM se destaca por su capacidad para separar clases con márgenes óptimos, mientras que RF ofrece robustez ante sobreajuste al combinar múltiples árboles de decisión.

Sin embargo, los modelos de aprendizaje profundo han superado estos resultados en escenarios de detección de comportamiento dinámico. Las *Convolutional Neural Networks* (CNN) han sido aplicadas con éxito al análisis de secuencias de *bytes* y a la identificación de patrones binarios de *malware* (Kolosnjaji et al., 2018), mientras que las *Long Short-Term Memory* (LSTM) destacan en la detección de comportamientos maliciosos en tiempo real en flujos de red (Pektaş & Acarman, 2020).

Por su parte, los *Autoencoders* y las *Graph Neural Networks* (GNN) son empleados para la detección de anomalías sin supervisión, aprenden representaciones de alta dimensión de tráfico normal y detectan desviaciones significativas (Wang et al., 2022).

Datos y entornos de experimentación

Las bases de datos más utilizadas en los estudios revisados fueron:

- EMBER Dataset (*Endgame Malware Benchmark for Research*).
- CICIDS2017 y CSE-CIC-IDS2018 (*datasets* de tráfico de red).
- VirusShare y Microsoft Malware Classification Challenge (BIG 2015) para análisis estático.
- NSL-KDD y TON_IoT *datasets*, relevantes en redes IoT y 5G.

Implementación y herramientas utilizadas

La revisión evidenció que los entornos más empleados para la experimentación son Python con TensorFlow, Keras, PyTorch y Scikit-learn, junto con plataformas de orquestación en la nube como Google Colab, AWS SageMaker y Azure ML. En investigaciones centradas en entornos críticos, los modelos fueron desplegados sobre redes SDN mediante mininet y OpenDaylight, con soporte para análisis en tiempo real (Al-Yaseen et al., 2020).

Limitaciones del estudio

Aunque el análisis ofrece una visión global, se identifican algunas limitaciones:

- Escasez de estudios centrados específicamente en redes troncales y sistemas de señalización.
- Falta de disponibilidad de *datasets* específicos de telecomunicaciones críticas.
- Dificultad para reproducir algunos resultados debido a la confidencialidad de los datos.
- Carencia de modelos explicables que permitan interpretar las decisiones de los algoritmos ante incidentes complejos.

Resultados y discusión

Los resultados de la revisión sistemática y del análisis comparativo de modelos de IA aplicados a la detección de *malware* en infraestructuras críticas de telecomunicaciones, evidencian la evolución sostenida de los métodos de detección hacia enfoques basados en aprendizaje profundo y detección contextual adaptativa. En esta sección se presentan los hallazgos principales, organizados según el tipo de modelo, su rendimiento, las métricas utilizadas y las implicaciones prácticas en el contexto de las redes de telecomunicaciones.

1. Rendimiento de los modelos de aprendizaje automático

Los modelos de aprendizaje automático clásico (SVM, Random Forest, Naïve Bayes, Decision Tree) siguen siendo competitivos en escenarios donde la detección de *malware* depende de características estáticas extraídas de archivos ejecutables o del análisis de cabeceras de tráfico. Según Permana et al. (2020), Random Forest alcanzó precisiones del 95,6 % en la clasificación de *malware*, mientras que SVM logró un rendimiento similar con tiempos de inferencia inferiores.

Sin embargo, su capacidad de detección disminuye ante muestras de *malware* polimórfico o metamórfico, así como en escenarios con tráfico cifrado o técnicas avanzadas de evasión. En el contexto de redes de telecomunicaciones críticas, donde los flujos son heterogéneos y de alta velocidad, los modelos clásicos tienden a generar una alta tasa de falsos positivos (FPR > 6 %), lo cual puede saturar los sistemas de alerta en centros de monitoreo (Cruz et al., 2021).

Por ello, las organizaciones del sector han comenzado a incorporar modelos híbridos, que combinan la capacidad interpretativa de los métodos clásicos con la potencia de generalización del aprendizaje profundo.

2. Eficacia de los modelos de aprendizaje profundo

El análisis reveló que los modelos basados en redes neuronales profundas (DNN, CNN, RNN, LSTM) ofrecen un rendimiento sustancialmente superior en la detección de *malware* dinámico y en la clasificación de patrones complejos de tráfico de red. En particular, las CNN han sido ampliamente utilizadas para representar el código binario del *malware* como una imagen matricial y detectar patrones maliciosos ocultos.

Kolosnjaji et al. (2018) demostraron que una CNN entrenada sobre imágenes binarias de ejecutables alcanzó una precisión del 98,4 % en la detección de familias de *malware* sin necesidad de desensamblar el código. Por su parte, LSTM y GRU resultan eficaces en la detección de comportamientos maliciosos persistentes al capturar relaciones temporales entre eventos en secuencias de tráfico (Pektaş & Acarman, 2020).

Los modelos Autoencoder destacan en la detección de anomalías

mediante reconstrucción de características. Al ser entrenados exclusivamente con tráfico legítimo, logran identificar desviaciones con tasas de precisión superiores al 97 %, lo que reduce de manera significativa los falsos positivos en entornos de alta disponibilidad (Wang et al., 2022).

Por otro lado, las GNN y los *Transformers* se posicionan como modelos emergentes de referencia, capaces de analizar dependencias estructurales entre nodos de red, identificar flujos coordinados de ataque y correlacionar alertas entre dominios distribuidos (Hussain et al., 2023).

La Tabla 1 resume las métricas promedio obtenidas de las principales investigaciones revisadas, comparando los modelos según precisión (*accuracy*), sensibilidad (*recall*), y tasa de falsos positivos (FPR).

Tabla 1. Rendimiento comparativo de modelos de IA en detección de *malware*

Tipo de modelo	Ejemplo de algoritmo	Precisión (%)	*Recall* (%)	FPR (%)
ML clásico	Random Forest	95.6	94.2	6.1
ML clásico	SVM	94.9	92.8	5.8
DL – CNN	CNN 2D en binarios	98.4	97.2	2.3
DL – LSTM	Secuencias de tráfico	97.6	96.9	3.1
DL – *Autoencoder*	Tráfico legítimo	97.2	95.8	2.6
GNN / *Transformer*	Topología de red	98.9	98.1	1.9

3. Aplicabilidad en infraestructuras de telecomunicaciones

En redes de telecomunicaciones, los flujos de datos se caracterizan por su alta dimensionalidad, diversidad de protocolos y baja latencia, lo que dificulta la detección de comportamientos anómalos en tiempo real. Los modelos de IA deben adaptarse a este contexto mediante técnicas de reducción de dimensionalidad y optimización de inferencia en el borde (*edge inference*).

Diversas investigaciones (Shafiq et al., 2020; Al-Yaseen et al., 2020) han implementado modelos CNN-LSTM integrados en entornos *Software Defined Networking* (SDN y han logrado detectar

ataques de tipo *botnet*, *ransomware* y *denial-of-service* en menos de 200 ms de tiempo de respuesta. En modelos mostraron una reducción del 35 % en la latencia de detección y una mejora del 40 % en la preparación respecto a métodos convencionales en escenarios experimentales que simularon infraestructuras de telecomunicaciones 5G.

Asimismo, la combinación de modelos Autoencoder con sistemas de detección basados en *flow analysis* permitió aislar comportamientos sospechosos en entornos de virtualización NFV (*Network Function Virtualization*), garantizando continuidad operativa y mitigación proactiva de amenazas (ENISA, 2023).

En el caso de las redes IoT vinculadas a telecomunicaciones, se evidenció que los modelos GNN son especialmente eficaces al representar la topología como un grafo dinámico, identificando relaciones entre dispositivos comprometidos y eventos coordinados (Zhang et al., 2022).

4. Ventajas y desafíos de la aplicación de IA

Entre las principales ventajas de los modelos de IA en la detección de *malware* destacan:

- Alta precisión en la identificación de amenazas conocidas y desconocidas.
- Adaptabilidad ante cambios en los patrones de ataque.
- Capacidad de aprendizaje continuo mediante actualización de pesos y *datasets*.
- Reducción de falsos positivos, optimizando los recursos de los centros de operaciones de seguridad (SOC).

Sin embargo, los desafíos son igualmente significativos:

- Explicabilidad limitada: la mayoría de los modelos de DL son cajas negras difíciles de interpretar (Goodman & Flaxman, 2017).
- Dependencia de datos etiquetados: la escasez de *datasets* específicos de telecomunicaciones limita el entrenamiento de modelos especializados.
- Vulnerabilidad ante ataques adversariales, que pueden engañar a los modelos de IA modificando levemente los patrones de entrada (Huang et al., 2021).

- Altos requerimientos computacionales, lo que dificulta la implementación en sistemas en tiempo real con recursos limitados.

5. Integración en sistemas de seguridad de telecomunicaciones

El despliegue de modelos de IA para detección de *malware* se ha integrado de manera progresiva en arquitecturas de seguridad existentes, como los *Security Information and Event Management* (SIEM) y los *Intrusion Detection Systems* (IDS) basados en IA. En entornos de telecomunicaciones críticas, esta integración requiere una orquestación inteligente capaz de correlacionar eventos provenientes de múltiples capas (física, lógica y de aplicación).

Sistemas como IBM QRadar, Splunk y Cisco SecureX han comenzado a incorporar módulos de detección basados en aprendizaje profundo para mejorar la correlación de alertas en redes 5G (CISA, 2022). Estos enfoques híbridos permiten la detección en tiempo real y la respuesta automatizada, lo que reduce en más del 60 % el tiempo medio de detección (MTTD).

Además, se están desarrollando *frameworks* federados de IA que posibilitan el entrenamiento de modelos de forma colaborativa entre operadores de telecomunicaciones sin compartir datos sensibles, lo que garantiza la privacidad mediante técnicas de aprendizaje federado (Yang et al., 2019). Este enfoque resulta prometedor para la defensa colectiva ante ciberataques de gran escala.

Conclusiones

El estudio realizado demuestra que la IA constituye una herramienta esencial para fortalecer la seguridad de las infraestructuras críticas de telecomunicaciones, y ofrece soluciones de detección de *malware* más precisas, adaptativas y escalables que los métodos tradicionales.

Los resultados comparativos indican que los modelos basados en aprendizaje profundo, especialmente las CNN, LSTM y GNN, superan de forma significativa a los enfoques clásicos de *Machine Learning* en términos de precisión, tasa de falsos positivos y capacidad de detección de amenazas desconocidas. Los modelos híbridos que combinan análisis estático, dinámico y de flujo de red han mostrado resultados

sobresalientes en entornos de alta criticidad, que alcanzan tasas de detección superiores al 98 %.

Asimismo, la integración de la IA con sistemas de seguridad existentes, como los IDS y SIEM, permite la correlación automatizada de eventos y la respuesta en tiempo real, lo que mejora la capacidad de detección y mitigación ante ciberataques complejos.

Sin embargo, persisten desafíos relevantes. Entre ellos se destacan la escasa disponibilidad de datos representativos de infraestructuras críticas, la interpretabilidad limitada de los modelos de aprendizaje profundo y la vulnerabilidad ante ataques adversariales, que pueden manipular las predicciones mediante perturbaciones mínimas en los datos de entrada. Además, la exigencia computacional de algunos modelos limita su despliegue en sistemas con recursos restringidos o en nodos perimetrales.

En consecuencia, se recomienda avanzar hacia el desarrollo de modelos explicables (XAI) que permitan comprender las decisiones de los algoritmos, la creación de *datasets* específicos del sector de telecomunicaciones y la adopción de enfoques federados que favorezcan el entrenamiento colaborativo entre operadores sin comprometer la privacidad de los datos.

La aplicación de IA en la detección de *malware* no solo representa una mejora técnica, sino un paso estratégico hacia la ciberresiliencia integral de las telecomunicaciones, garantizando la continuidad operativa de los servicios que sustentan la economía digital y la seguridad nacional.

Referencias bibliográficas

Alasmary, W., Alasmary, M., & Alhaidari, F. (2022). AI-based Malware Detection and Classification: A Survey. Computers & Security, 121, 102856. *Elsevier*. https://doi.org/10.1016/j.cose.2022.102856

Al-Yaseen, W. L., Othman, Z. A., & Nazri, M. Z. (2020). Hybrid Intelligent Intrusion Detection System Using Machine Learning Techniques. *Journal of Information Security and Applications*, 55, 102607. https://doi.org/10.1016/j.jisa.2020.102607

CISA – *Cybersecurity and Infrastructure Security Agency. (2022). Securing Communications Infrastructure: Guidance for Critical Networks*. U.S. Department of Homeland Security. Disponible en: https://www.cisa.gov

Cruz, A., Márquez, J., & Ramírez, D. (2021). Machine Learning Models for Malware Detection in Telecommunication Networks. *IEEE Access, 9*, 178543–178557. https://doi.org/10.1109/ACCESS.2021.3130201

ENISA – European Union Agency for Cybersecurity. (2023). *Threat Landscape for the Telecom Sector 2023. European Commission*. Disponible en: https://www.enisa.europa.eu

García-Teodoro, P., Maciá-Fernández, G., Díaz-Verdejo, J., & Vázquez, E. (2023). Advanced Intrusion Detection in Telecommunications: AI-driven Approaches. *Computer Networks, 235*, 109000. https://doi.org/10.1016/j.comnet.2023.109000

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". *AI Magazine, 38(3), 50–57*. https://doi.org/10.1609/aimag.v38i3.2741

Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B. I. P., & Tygar, J. D. (2021). Adversarial Machine Learning. *Communications of the ACM, 64(7), 64–73*. https://doi.org/10.1145/3434228

Hussain, F., Abbas, G., Fatima, R., & Wang, H. (2023). Graph Neural Networks for Network Threat Detection in 5G and IoT Environments. *IEEE Internet of Things Journal, 10(8)*, 7254–7267. https://doi.org/10.1109/JIOT.2023.3241073

Kolosnjaji, B., Zarras, A., Webster, G., & Eckert, C. (2018). *Deep Learning for Classification of Malware System Call Sequences. In Proceedings of the 2018 IEEE International Conference on Big Data (pp. 3787–3796)*. IEEE. https://doi.org/10.1109/BigData.2018.8621968

Pektaş, A., & Acarman, T. (2020). Malware Classification Based on Deep Learning Architectures for Android. *Neural Computing and Applications*, 32(3), 817–833. https://doi.org/10.1007/s00521-018-3930-1

Permana, A., Pradana, F., & Utami, R. (2020). *Comparative Analysis of Machine Learning Algorithms for Malware Detection. International Journal of Computer Applications, 175*(13), 14–22. https://doi.org/10.5120/ijca2020920431

Sgandurra, D., & Lupu, E. C. (2020). Automated Malware Classification and Analysis Using Machine Learning. *Computers & Security*, 95, 101867. https://doi.org/10.1016/j.cose.2020.101867

Shafiq, M., Tian, Z., Bashir, A. K., Du, X., & Guizani, M. (2020). Corrauc: A Deep Learning Model for Anomaly Detection in 5G Telecom Networks. *IEEE Transactions on Network Science and Engineering*, 7(4), 2348–2362. https://doi.org/10.1109/TNSE.2020.3025241

Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of Machine Learning Techniques for Malware Analysis. *Computers & Security*, *81*, 123–147. https://doi.org/10.1016/j.cose.2018.11.001

Wang, J., Zhang, H., & Zhu, X. (2022). Deep Autoencoder-Based Network Anomaly Detection for Critical Infrastructure. *Expert Systems with Applications, 201*, 117138. https://doi.org/10.1016/j.eswa.2022.117138

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated Machine Learning: Concept and Applications. ACM Transactions on Intelligent Systems and Technology, 10(2), 12. https://doi.org/10.1145/3298981

Zhang, Y., Pan, S., & Chang, X. (2022). Graph Representation Learning for Cybersecurity: *A Review. IEEE Transactions on Neural Networks and Learning Systems*, 33(5), 1805–1822. https://doi.org/10.1109/TNNLS.2021.3114520

INVESTIGACIÓN

Plataforma integrada para la firma digital en el contexto de la empresa XETID

Integrated platform for digital signatures in the context of the company XETID

Dr.C. Huber Martínez Rodríguez[*], Ing. Yosvel Dupeirón Porra[2]

Recibido: 11/2025 | Aceptado: 11/2025 | Publicado: 12/2025

Resumen

En el proceso de informatización de la sociedad en los tiempos actuales, es imprescindible tener a disposición una Infraestructura de Clave Pública (PKI) en integración con un servicio seguro de firma digital disponible para su uso por la sociedad y el sector corporativo. La firma digital garantiza la integridad y la asociación del firmante con el documento digital firmado, y asegura así la autenticidad del este con fines legales. El presente trabajo propone una plataforma, que integra diferentes componentes, con el objetivo de brindar un servicio de firma digital seguro desde la Empresa de Tecnologías de Información para la Defensa (XETID) para la sociedad cubana en general. En este sentido intervienen, una infraestructura de llave pública, un servidor de firma digital, un servidor de sellado de tiempo y un servicio seguro para el manejo de claves y operaciones de criptográfica. También, se proponen tipologías de herramientas para el disfrute del servicio de firma digital.

[*] Empresa de Tecnologías de la Información para la Defensa, XETID, La Habana, Cuba. huber@xetid.cu

Universidad Máximo Gómez Báez de Ciego de Ávila, Cuba. huber@unica.cu

[2] Empresa de Tecnologías de la Información para la Defensa, XETID, La Habana, Cuba. ydupeiron@xetid.cu

En el desarrollo del trabajo se describen cada uno de los componentes, sus funciones esenciales y el tratamiento a las claves, con el propósito de lograr un servicio de firma digital robusto y seguro. Así mismo, se describen algunas herramientas esenciales para posteriores desarrollos con el fin de lograr el objetivo deseado.

Palabras clave: infraestructura de clave pública, servicio de sellado de tiempo, servidor de firma digital, protección segura de claves

Abstract

In the current process of digitizing society, it is essential to have a Public Key Infrastructure (PKI) integrated with a secure digital signature service available for use by society and the corporate sector. A digital signature ensures the integrity of and the signer's association with the signed digital document, thereby guaranteeing its authenticity for legal purposes. This paper proposes an integrated platform from the XETID Company that aims to provide Cuban society with a secure digital signature service at large. The solution involves a public key infrastructure, a digital signature server, a timestamp server, and a secure service for managing passwords and performing cryptographic operations. Additionally, tool typologies for utilizing the digital signature service are proposed. The work describes each component, its essential function, and the handling of passwords necessary to achieve a robust and secure digital signature service. Likewise, it outlines essential tools for further development so as to accomplish the desired objective.

Keywords: *public key infrastructure, timestamp service, digital signature server, secure password protection*

Introducción

La firma digital se ha convertido en una herramienta cada vez más demandada por las instituciones y la sociedad cubana en general. En el proceso de informatización de los servicios, la firma digital es indispensable y, con ella, se garantiza la integridad y la asociación con el firmante de un documento digital, asegurando así la autenticidad de este y su validez legal. Pero no se trata solo de validar

documentos, la firma digital es la llave que permite reducir papeles, agilizar procesos y garantizar trazabilidad legal en cada documento.

En 2016, Cuba inició la formalización de las firmas digitales mediante la adopción de la infraestructura de clave pública (PKI), según los lineamientos marcados por la *Resolución 2 del Ministerio del Interior*, luego derogada por el *Decreto-Ley 79/2023*[1] y su reglamento el *Decreto 106/2024*[2]. Esta infraestructura garantiza la emisión de certificados digitales, que son documentos electrónicos que acreditan la identidad de una persona o entidad por lo que podrían tener asociada un par de claves (**clave pública y clave privada**) que les permite asociar su identidad en el mundo digital.

En 2018, el *Decreto-Ley* 370 sobre la informatización de la sociedad en Cuba, reconoció de forma legal la validez de los documentos firmados electrónicamente en el país. En 2022, el Ministerio del Interior (MININT) y el Consejo de Ministros crearon reglas más claras para el uso de firmas digitales en órganos gubernamentales, el Banco Central de Cuba y agencias locales. Ya en la actualidad el Decreto Ley 79/2023 y su reglamento el Decreto 106/2024 sientan las bases jurídicas para las infraestructuras de clave pública en Cuba.

El uso de una PKI como componente individual no garantiza un despliegue adecuado de la firma digital, o sea, solo avala la emisión de certificados digitales que están asociados a la identidad. El manejo de las claves es de suma importancia cuando se diseña una plataforma para la firma digital. En este escenario se trabaja con criptografía asimétrica (Menezes, 1996; Cohen, 2006), así cada persona o entidad contará con un par de claves, es decir, una **clave pública** que como su nombre lo indica, es de conocimiento público y una **clave privada** que debe estar muy bien resguardada. Existen estándares tanto de *software* PKCS#12 (.p12, .pfx, .jks) como de *hardware* (módulos

[1] Decreto Ley 79/2023 "Sobre el desarrollo, la aplicación y uso de los dispositivos de protección criptográfica y servicios en la esfera de la criptografía en la República de Cuba" (GOC-2024-527-O89)

[2] Decreto 106/2024 Reglamento del Decreto Ley 79 "Sobre el desarrollo, la aplicación y uso de los dispositivos de protección criptográfica y servicios en la esfera de la criptografía en la República de Cuba", de 26 de octubre de 2023 (GOC-2024-528-O89)

de seguridad de *hardware*, HSM) que permiten, mediante algoritmos criptográficos, asegurar las claves (Gitlan, 2025). La clave pública es usada en el proceso de **cifrado** y la clave privada en el proceso de **descifrado** de la información. Es esencial garantizar que las claves privadas no viajen por canales inseguros o espacios que faciliten la captura de estas por intrusos, en infraestructuras críticas es importante establecer políticas y procedimientos para el manejo de claves que sean robustos y con altos niveles de seguridad. Kuzminykh (2021) hace un estudio detallado de varias herramientas para el manejo de claves (*KMS, Key Management Service*) donde se definen criterios como *software* libre o de pago, almacenamiento seguro, soporte para HSM, uso de *logs* de auditorías, entre otros.

Además, son necesarias herramientas para llevar a cabo la firma de forma transparente y sencilla. Estas herramientas pueden ser diversas, es decir, una APK para celulares, una app para escritorio u otro tipo de aplicación siempre que tenga garantías de seguridad. Por último es esencial un servidor de firma digital, este debe estar dotado de buenas funcionalidades, debe interactuar con un servicio de sellado de tiempo para alcanzar la validez de un documento firmado aún cuando el certificado haya caducado y en ese caso, los sellos de tiempo son esenciales. En el entorno cubano, diferentes empresas hacen un gran esfuerzo para el desarrollo de la firma digital, tanto para fines comerciales como para su desarrollo empresarial. Algunos ejemplos son: XETID, Softel, Datys, Tecnomática y Segurmática que tienen PKI certificadas por el MININT (Hernández, 2024). Cuando se revisa en cada propuesta de infraestructura de clave pública, la integración de todos los elementos antes mencionados no es alcanzado en su totalidad.

Basado en las razones anteriores, en este trabajo se traza como objetivo: proponer una plataforma integrada y segura para la firma digital, en el entorno de la empresa XETID. Con ese propósito se diseña la interacción general de la plataforma y se plantea para cada uno de sus componentes los diseños y herramientas, así como las funcionalidades deseadas y los estándares de seguridad a alcanzar.

Materiales y métodos

La presente propuesta está basada en la experiencia adquirida en el desarrollo y despliegue de Rinho PKI (desarrollado por: Instituto de Criptografía de la Universidad de la Habana) en la infraestructura de la empresa XETID, basada principalmente en criptografía asimétrica con curvas elípticas (Menezes, 2006) y en armonía con los estándares internacionales en este campo. Se debe mencionar además que el componente principal, la PKI, ya tiene un grado de madurez importante y que se cuenta con todos los recursos e infraestructuras para emprender la propuesta aquí expuesta.

Para potenciar la informatización de la sociedad cubana y optimizar los servicios que la empresa XETID ofrece a terceros, se identifica la necesidad de desarrollar nuevos proyectos que integren el ecosistema de la firma digital. Cabe aclarar, que las funcionalidades propuestas para cada componente son mínimas y pueden ampliarse según las necesidades que surjan durante el desarrollo. No obstante, los requisitos aquí expuestos para el tratamiento de las claves se mantienen estrictos. El presente trabajo se centra en los protocolos que garantizan una interacción segura entre los componentes, sin profundizar en los detalles funcionales internos de cada uno.

La propuesta cuenta con seis componentes fundamentales:

1. **Infraestructura de clave pública**: una PKI incluye las políticas, las funciones, el *hardware*, el *software* y los procedimientos necesarios para crear, administrar, distribuir, usar, almacenar y revocar certificados digitales. Tiene como objetivo facilitar la transferencia electrónica segura de información para una serie de actividades en red. Como, por ejemplo: el comercio electrónico, la banca y el correo electrónico confidencial, entre otros. Intervienen en el proceso:
   - **Certificados digitales**: credenciales electrónicas que vinculan la identidad del titular del certificado a un par de claves que pueden utilizarse para cifrar y firmar información.
   - **Autoridad de certificados (CA)**: una entidad de confianza, autorizada y certificada por el MININT, que emite certificados digi-

tales. La CA gestiona todos los aspectos de los certificados PKI, incluido su ciclo de vida.

- **Autoridad de registro (RA)**: se encarga de aceptar las solicitudes de certificados y autenticar a la persona u organización que las presenta.
- **Autoridad de validación (VA)**: se encarga del proceso de validación de certificados a partir de servicios de OCSP (*Online Certificate Status Protocol*) y CRL (*Certificate Revocation List*).
- **Aplicación web de gestión centralizada**: es una web a partir de la cual los usuarios pueden solicitar y gestionar sus certificados digitales.

2. **XFirma APK**: es una aplicación para celulares que permite, la firma del lado del usuario y avala el manejo privado de sus claves en el entorno de su dispositivo, es decir, este usaría su archivo de claves .p12 para el proceso de firma o podría obtener un certificado desde la CA a partir de una solicitud CSR construida por la misma aplicación XFirma, cuestión más recomendable. Posibilita, además, la validación de certificados y documentos firmados por la APK o por terceros. Este tipo de aplicación tiene potencial para ampliar sus funcionalidades libremente, lo que garantiza el cifrado de documentos, compartir documentos entre usuarios bajo ciertos requerimientos, el envío de documentos cifrados por correo, previo intercambio de claves, entre otras funcionalidades.
3. **Servidor de firma**: se encarga de gestionar la firma digital de documentos en formato .pdf, .xml, códigos, entre otros, y se usan fundamentalmente estándares europeos para la firma digital (eIDAS, Reglamento [9/10/2014]). Es factible su uso en entornos corporativos o para el desarrollo del gobierno electrónico. Debe ser flexible en las interfaces para clientes certificando su uso mediante API Rest, servicios web, Internet, o a través del uso de una herramienta de firma integrada al servidor para el cliente. Para la firma de un documento, el servidor gestiona el aspecto de la firma, el proceso de conformación del documento firmado y además la integración en este del sello de tiempo.

4. **Servidor TSA**: un servidor de sellado de tiempo RFC3161 cumple una función esencial en la protección a largo plazo de los registros de datos. Proporciona la prueba de que los datos existían en un momento determinado y que no han sufrido cambios, ni siquiera un solo bit binario, desde su certificación y sellado de tiempo. El servicio de sellado de tiempo es esencial en una PKI. Es recomendable la renovación por lo menos anual de los certificados de firma usado por el servidor TSA.

5. **Servicio de manejo de claves (SMC) para la firma digital**: en el funcionamiento de un servidor de firma digital es esencial hacer un manejo adecuado de las claves y su resguardo, así en la presente propuesta se habilita este componente donde el usuario o entidad se registra y solicita la generación de sus claves para la firma digital. Este servicio puede tener las siguientes funcionalidades: gestión del ciclo de vida completa, incluyendo su creación, habilitación, deshabilitación, eliminación, rotación, modificación y la asignación de alias; gestión de claves de datos, incluyendo la creación, el cifrado y el descifrado de claves de datos; creación y gestión de claves simétricas; entre otras funciones.

A partir del uso del servicio de solicitud de certificados usando CSR de la CA, se genera localmente la clave privada y pública en el SMC y se protege la llave privada, se crea la solicitud CSR y se solicita el certificado correspondiente firmado por la CA. De esta forma, se garantiza la existencia de claves para el usuario o la entidad en la SMC. Además, las claves están altamente protegidas. Esta protección es recomendada con el uso de un HSM. Es importante señalar que en este entorno se hacen los procesos criptográficos necesarios con las claves aquí almacenadas (firma, cifrado, entre otros) y se utilizan solicitudes con el estándar PKCS#11, así se garantiza que estas claves no salen del entorno seguro.

6. **XFirma Corp.**: es una aplicación web que gestiona la firma digital para todos sus usuarios o entidades registradas, usa una o varias de las interfaces del servidor de firma para gestionar la firma de documentos, XML o código. En este entorno, no es seguro el uso de los .p12 de los usuarios, dado que tendrían que viajar hasta el

servidor de XFirma Corp. además de confiar en una entidad no diseñada para la protección de claves. Si la intención del usuario es firmar con su .p12 entonces lo más seguro es usar XFirma APK. Esta aplicación web también permite incorporar funcionalidades relacionadas con el intercambio de documentos firmados entre usuarios mediante el intercambio previo de claves o el uso de un servicio de correo seguro.

El uso de un servidor de identidad para la integración de los componentes en armonía con sus funcionalidades es de suma importancia. Se lograría así obtener un token de acceso único para los componentes relacionados.

Para los diseños aquí descritos se empleó el aplicativo *draw.io* que permitió de forma libre definir las interacciones en cada uno de los componentes propuestos.

Resultados y discusión

La Figura 1 muestra el diseño sin detalles funcionales de los componentes que intervienen en la propuesta antes comentada y su interrelación.

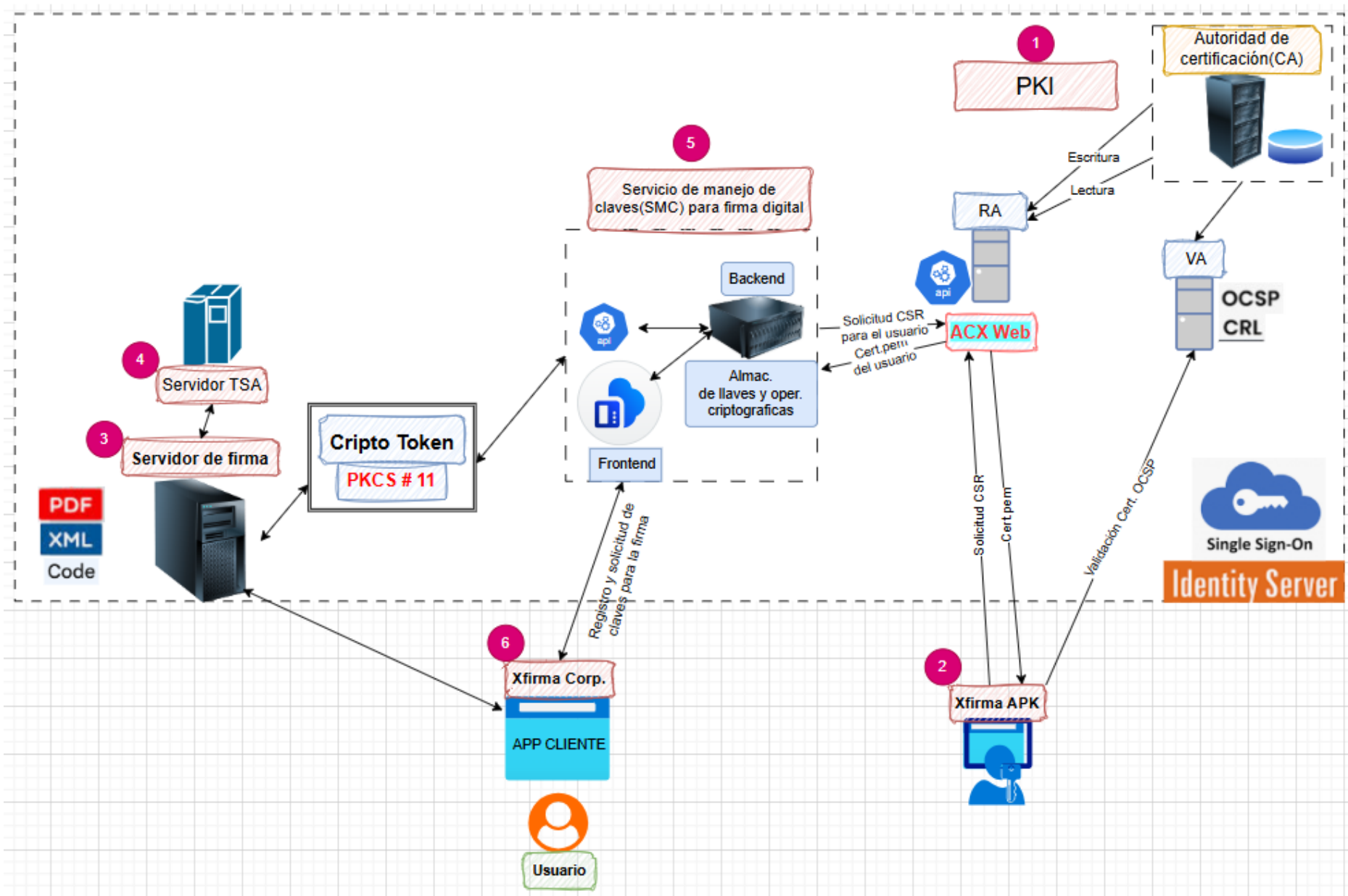

Figura 1. Desglose numerado de los componentes y su interacción dentro del ecosistema de la firma digital

Desglose de interacción seguras entre componentes y funciones relacionadas con la firma digital

1. **Interacción XFirma APK o app con ACX**: al utilizar la librería Bouncy Castle con un generador de número aleatorio propio, se obtienen claves pública y privada en el dispositivo del usuario. Luego debe mantenerse en resguardo la clave privada utilizando archivos cifrados y, después, se genera la solicitud de certificados CSR y se envía a la RA a través de la interfaz de usuario definida. Luego descarga el certificado dado en un archivo .pem y firmado por la CA. A continuación se debe construir el contenedor de claves mediante un archivo .p12 que será guardado en un lugar seguro por la aplicación. Estas claves estarán asociadas al usuario que previamente se registró en la aplicación mediante un formulario de registro adaptado a sus necesidades y requerimientos en su interacción con la PKI. Por último se puede firmar un documento del lado del usuario en su dispositivo, la aplicación valida la firma y además el certificado usado es validado contra el OCSP o usando CRL.
2. **Interacción de XFirma Corp. con el SMC**: el cliente se registra y crea su perfil en el SMC, solicita además su correspondiente par de claves para la firma digital. En este caso, el servidor se encarga, con los datos del usuario, de hacer la solicitud con un CSR de su certificado a la CA. Con el certificado y la clave privada debidamente protegida, se crea y almacena su contenedor .p12 en el SMC. De esta forma ya el usuario tendrá su registro y las claves para la firma digital en el servidor SMC. Es importante hacer notar que el SMC puede tener muchas funcionalidades que no es objetivo discutir en esta propuesta.
3. **Interacción de XFirma Corp. con el Servidor de Firma**: el usuario solicita la firma de un documento en su perfil en XFirma Corp. Esta aplicación web emplea la interfaz de usuario establecida envía al Servidor de Firma, el IDu del usuario, el IDc del certificado y el documento a firmar (doc). Se supone que el usuario ya tiene creado su perfil de firma en el Servidor de Firma con el aspecto para la firma configurado.

4. **Interacción del Servidor de Firma con el SMC**: en este caso, el Servidor de Firma calcula el *hash* del documento (**h=hash(doc)**) enviado por XFirma Corp. y envía para el SMC el IDu, IDc y el **valor *hash* h** del documento usando el estándar PKCS #11 con la orden además de firmar el valor *hash* h. El servidor SMC, por su lado, firma el valor *hash* h del documento con las claves y el usuario especificado, finalmente devuelve ese *hash* firmado al Servidor de Firma.
5. **Interacción del Servidor de Firma con el Servidor TSA**: el Servidor de Firma envía al Servidor TSA el *hash* firmado para el sellado de tiempo, el Servidor TSA agrega el sello de tiempo al *hash* firmado y firma este con sus claves. Es agregado además el certificado del TSA y la cadena de certificación obteniendo así el **Timestamp Token** que es enviado al Servidor de Firma.
6. **Interacción Servidor de Firma con XFirma Corp.**: ahora el Servidor de Firma construye el documento firmado y lo envía de regreso **XFirma Corp**. Así el usuario obtiene el documento firmado sin la necesidad de que sus claves estén viajando por la infraestructura.

Herramientas recomendadas en los servidores

- **Infraestructura de clave pública**: es muy usada en los entornos empresariales EJBCA de la empresa KEYFACTOR con una larga experiencia en desarrollos para la seguridad de la información. Esta herramienta tiene una versión *community* y de código abierto que permite el desarrollo sobre esta versión. Se programa en Java y se sustenta en la librería criptográfica Bouncy Castle. La versión *Enterprise* de EJBCA tiene más funcionalidades, pero es de pago. Para un entorno de producción basado en EJBCA se promueve el uso en el contexto cubano de Rhino PKI, desarrollado y comercializado por el Instituto de Criptografía de la Universidad de La Habana.
- **XFirma APK**: puede desarrollarse una APK para Android usando una librería criptográfica en correspondencia con los tipos de documentos a firmar. Existen varias que son libres y, si se sigue el desarrollo de la PKI hay una forma de utilizar en Android la librería criptográfica Bouncy Castle teniendo así una librería más completa.
- **XFirma Corp.**: Es un desarrollo web que puede integrar diferentes tecnologías y lenguajes, en dependencia de las necesidades.

Bouncy Castle es una librería desarrollada para los lenguajes Java y C#, por tanto, al usar JavaScript o Python para llevar a cabo el desarrollo web cambiarían las librerías criptográficas a usar. También es importante tener en cuenta que se necesitan librerías que hagan tratamiento de la firma digital, validación y manipulación de certificados digitales con el estándar X509. Por ejemplo, si se decide usar el lenguaje JavaScript, en este caso una librería bastante completa es **node-forge** que es similar a Bouncy Castle. Como propuesta completa se tiene:

Backend: Node.js + Express/Fastify, TypeScript (opcional, pero recomendado), Jest para pruebas.

Para criptografía: node-forge / libsodium; bcrypt para passwords; jsonwebtoken para autenticación.

Para base de datos: PostgreSQL con pgcrypto o MongoDB con field-level encryption.

Para frontend: React/Vue/Angular, Web Crypto API para operaciones del cliente.

- **Servidor de Firma**: Se propone el uso de la versión *community* del servidor SignServer[4] de la empresa KEYFACTOR[5] que es de código abierto y sobre esa versión desarrollar un servidor de firma en correspondencia a las necesidades. En la Figura 2 se muestra la interacción del usuario con el servidor y las interfaces disponibles para aplicaciones clientes o para usuario con el fin de disfrutar de los servicios del Servidor de Firma.

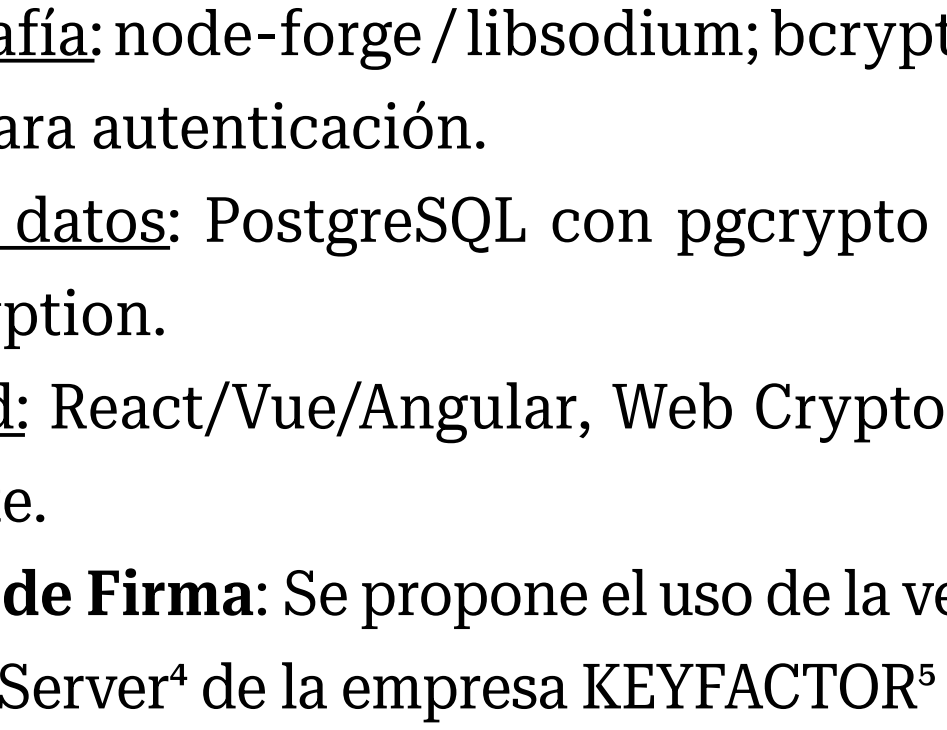

Figura 2. Representación de las funciones del Servidor de Firma

[4] https://www.signserver.org/

[5] https://www.keyfactor.com/

- **Servidor TSA**: Existe un ecosistema de herramientas que se pueden usar como servicio TSA para la firma digital o desarrollar sobre ellas. Por ejemplo: <www.freetsa.org>, <http://timestamp.digicert.com>, OpenTSA, entre otras opciones. No obstante, la instauración de un servidor TSA para producción requiere de desarrollos en correspondencia a la demanda del servicio. En el caso empresarial, para producción puede ser recomendable usar el servidor TSA incluido en la plataforma EJBCA que permite el trabajo con imágenes Docker, balance de carga y adaptación a la demanda (Figura 3).

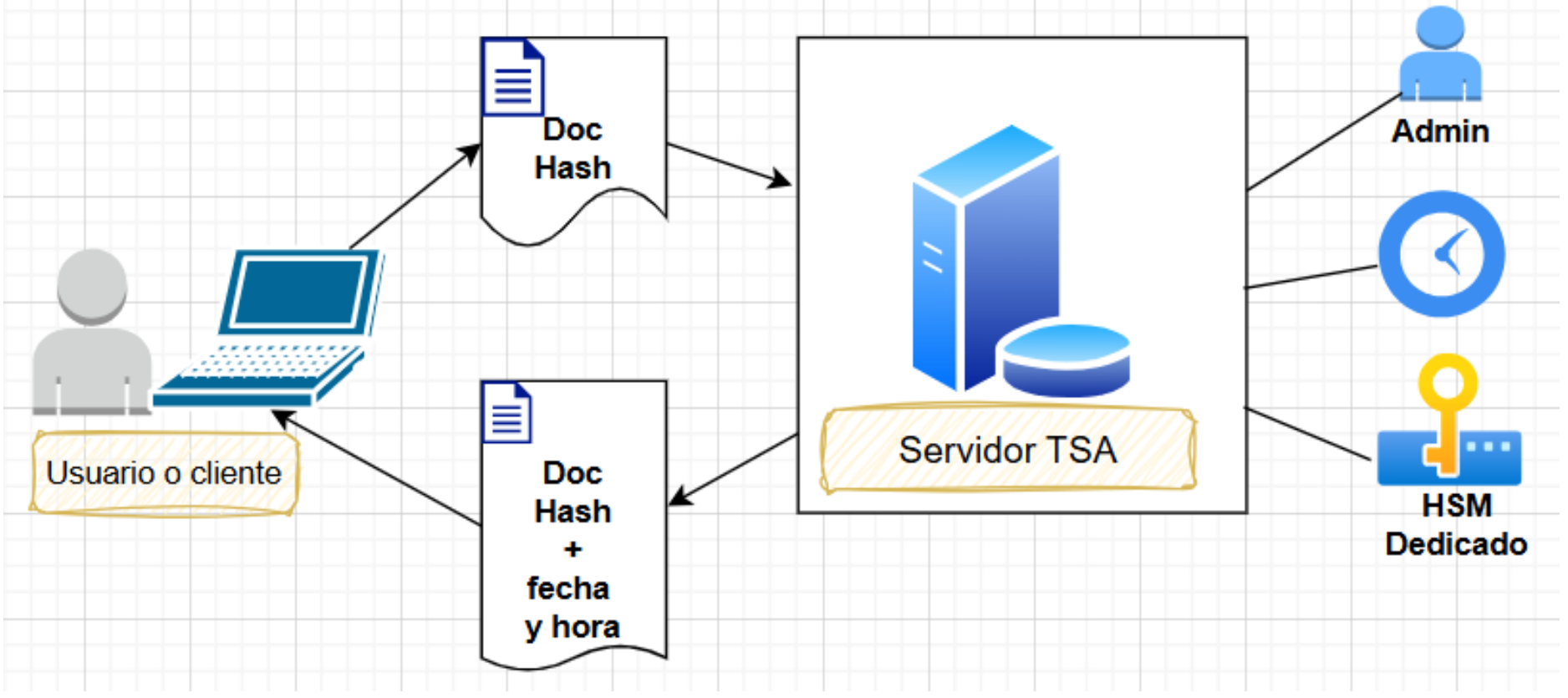

Figura 3. Estructura organizativa recomendada para un Servidor TSA

- **Servicio de Manejo de Claves (SMC) para la firma digital**: para los servicios de manejo de claves existen muchas opciones en la nube con este fin. Kuzminykh (2021) hace un análisis de una amplia gama de herramientas que realizan esta función con criterios bien delimitados y permiten tomar una decisión al momento de elegir una de ellas. En el entorno de la empresa XETID es recomendable llevar a cabo un desarrollo propio usando algunos de los servicios libres con estos fines como SoftHSMV2, que se integra fácilmente con el SignServer de KEYFACTOR y facilita la interacción entre el servidor de firma y el Servidor SMC usando el estándar PKCS#11. Es importante hacer notar que es una salida a la ausencia de una HSM profesional, pero no es la opción recomendada. No obstante, si se extrema la seguridad en la infraestructura y el servicio, podría ser una opción viable.

El desarrollo de una plataforma de esta envergadura requiere dedicación y entrega de todo un equipo de trabajo para el logro adecuado de una integración segura y amigable al usuario final. El trabajo modular de cada componente es esencial con vistas al logro de los avances deseados.

Conclusiones

A través de la investigación, se constata la necesidad de una plataforma con potencial para contribuir a la informatización de la sociedad cubana con el uso de la firma digital. La diversidad de herramientas existente de código abierto para cada uno de los componentes descritos obliga a realizar una revisión profunda para evaluar su factibilidad en el despliegue a nivel de producción. En todos los entornos seguros son recomendables el uso de módulos HSM dedicados, pero dependiendo del contexto podría sustituirse por uno similar a nivel de *software*, lo que garantiza fortalecer la seguridad con medidas adicionales, aunque no es lo recomendable. Es de suma importancia, que el servidor de firma tenga diferentes opciones de interfaz de usuarios con la finalidad de diversificar los tipos de servicios a usuarios o clientes. En los proyectos futuros de la empresa se promueve dar continuidad a los desarrollos actuales integrando los servicios descritos en cada uno de los componentes de la plataforma propuesta.

Referencias bibliográficas

Cohen, H., Frey, G., Avanzi, R., Doche, C., Lange, T., Nguyen, K. y Vercauteren, F. (2006). *Handbook of elliptic and hyperelliptic curve cryptography*. Chapman and Hall/CRC.

Gitlin, D. (2025, 27 de octubre). *Cómo almacenar claves privadas de forma segura*. SSL Dragon. https://www.ssldragon.com/es/blog/mejores-practicas-almacenar-clave-privada/.

Hernández, A. (2024, 21 de febrero). *La firma digital, una herramienta clave para la informatización de la sociedad cubana*. Cubadebate. http://www.cubadebate.cu/especiales/2024/02/21/la-firma-digital-una-herramienta-clave-para-la-informatizacion-de-la-sociedad-cubana/.

Kuzminykh, L., Ghita, B. V. y Shiaeles, S. (2021). *Comparative analysis of cryptographic key management systems*. arXiv. https://doi.org/10.48550/arXiv.2109.09905.

Menezes, A. J., van Oorschot, P. C., y Vanstone, S. A. (1996). *Handbook of applied cryptography*. CRC Press.

CRONOLOGÍA

Impacto de las nuevas tecnologías en la ciberseguridad

The impact of new technologies on cybersecurity

M. Sc. Reniel Carvajal Alfonso[*], M. Sc. Dianelys Alvarez González[2]

Recibido: 11/2025| Aceptado: 11/2025| | Publicado: 12/2025

Resumen

El desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC) ha posibilitado, junto al avance de las redes de comunicaciones 5G y 6G, la evolución del mercado de las *Fintech*. Dicho mercado facilita el intercambio de información, pero, a su vez, ha propiciado que ciberdelincuentes roben a los gobiernos y a los TELCOS información sensible con el objetivo de causar la inoperancia de estos y obtener, por medio de chantaje, financiamientos. Esta investigación analiza el creciente impacto de la ciberseguridad en el sector de las telecomunicaciones. El estudio describe amenazas específicas como los ataques de denegación de servicio distribuidos (DDoS), los troyanos bancarios, las vulnerabilidades *zero-day* y *malware* impulsado por inteligencia artificial. Como resultado se evidencia el impacto de las nuevas tecnologías como IA, Internet de las Cosas (IoT) y el comercio electrónico han incrementado la vulnerabilidad, así como la interrelación entre los organismos reguladores, empresas, gobiernos a través de marcos regulatorios y la capacitación constante de los usuarios como elemento crucial para hacer frente a estas amenazas.

[*] Empresa de Telecomunicaciones de Cuba S.A., ETECSA. La Habana 11300. Cuba. reniel.carvajal@etecsa.cu

[2] Empresa de Telecomunicaciones de Cuba S.A., ETECSA. La Habana 11300. Cuba. dianelys.alvarez@etecsa.cu

Palabras clave: inteligencia artificial, ciberataques, TELCOS, ciberseguridad, ciberdelincuentes

Abstract

Advances in Information and Communication Technologies (ICT) together with those of 5G and 6G communications networks have facilitated the exchange of information and enabled the evolution of the Fintech market. However, cybercriminals have also targeted these technologies with the firm determination to render TELCOs and governments inoperative and obtain financial gain through blackmails and the theft of sensitive information as a result of these cyberattacks. This study focuses on analyzing the increasing impact of cybersecurity on telecommunications industry. Threats such as Distributed Denial of Service (DDoS) attacks, banking Trojans, zero-day vulnerabilities, and AI-driven malware are described. As a result, it is clear that new technologies such as Artificial Intelligence (AI), the Internet of Things (IoT), and e-commerce have increased vulnerability and the interrelation between regulatory bodies, companies, and governments. This is evident through regulatory frameworks and the constant training of users, which are crucial elements in addressing these threats.

Keywords: *Artificial Intelligence, cyberattacks, TELCOs, cybersecurity, cybercriminals*

Introducción

La ciberseguridad emerge como una disciplina esencial, derivada del avance en infraestructura de red, *hardware*, *software* y del incremento en la interconexión y digitalización de servicios estratégicos como la Bancarización y las soluciones *Fintech*. Si bien estos procesos generan beneficios significativos, también exponen a las organizaciones a un creciente riesgo y las convierten en blancos potenciales de ataques maliciosos. El objetivo de tales ataques es el robo de información, el daño de activos digitales o la interrupción de la continuidad operativa, con el fin último de generar desestabilización, pérdidas económicas y deterioro de la confianza de los clientes.

La Oficina de las Naciones Unidas contra las Drogas y el Delito considera que la ciberdelincuencia es un concepto muy complejo

que engloba una variedad de actividades ilícitas. En este sentido, precisa que la *ciberdelincuencia* es un acto que infringe la ley y que se comete usando las TIC para atacar las redes, sistemas, datos, sitios web, entre otros (Ministerio de Relaciones Exteriores de Cuba, 2021). Varios autores defienden el ciberdelito, teniendo en cuenta lo anterior, como aquel delito que se comete mediante las Tecnología de la Información y las Comunicación (TIC), contra individuos, empresas o gobiernos. Por lo tanto, el ciberdelito constituye una acción antijurídica que está contemplada en el derecho penal (Benito, 2023).

De acuerdo a los expertos de la Asociación de Auditoría y Control de Sistemas de Información (ISACA), la ciberseguridad se define como una capa de protección para los archivos de información. También, para referirse a la ciberseguridad se utiliza el término seguridad informática o seguridad de la información electrónica (Infosecurity Mexico, 2025).

Es un hecho que la expansión de la conectividad conlleva una necesidad urgente de proteger no solo los datos personales de los usuarios, sino también la infraestructura crítica de las empresas. En este contexto, resulta imprescindible fomentar la interoperabilidad entre operadores de telecomunicaciones (TELCOS), agencias reguladoras y gobiernos. El objetivo es implementar estrategias conjuntas que permitan prevenir, detectar y mitigar ciberamenazas. Asimismo, es esencial fortalecer las capacidades institucionales y ciudadanas mediante la formación continua, esta medida garantiza una protección integral frente a los riesgos emergentes del entorno digital.

Materiales y métodos

Esta investigación adopta un enfoque metodológico de carácter mixto con preponderancia cualitativa. Se emplearon los métodos **inductivo-deductivo** y **analítico-sintético** para el examen de la información. La técnica de análisis documental fue fundamental para la selección y evaluación crítica de las fuentes bibliográficas, lo que permitió una profundización sistemática en el tema de estudio.

Panorama de la Ciberseguridad a nivel global

Según el Fondo Monetario Internacional (FMI) estima que la ciberdelincuencia costará a nivel global 23 billones de dólares para el año 2027. En la región de Latinoamérica se ha experimentado un aumento de 108 % de ciberataques a operadores de telecomunicaciones, respecto al año anterior, con más de 2,600 incidentes semanales por organización. México, Colombia y Brasil encabezan los reportes en volumen y sofisticación de amenazas. El costo promedio de violación de datos de 4,88 millones de dólares en 2024. Ese mismo año, el *ransomware* se consolidó como la principal amenaza, tanto a nivel global como en Latinoamérica y el Caribe, al representar un 38 % del total de todas las ciberamenazas registradas. Los principales sectores de ataque de esta amenaza se centraron en comercio mayorista y servicios, salud y construcción. Los principales *ransomware* con mayor impacto en el área de Latinoamérica fueron RansomHub, LockBit y Akira. En la Figura 1. se muestra un resumen del impacto global de esta amenaza.

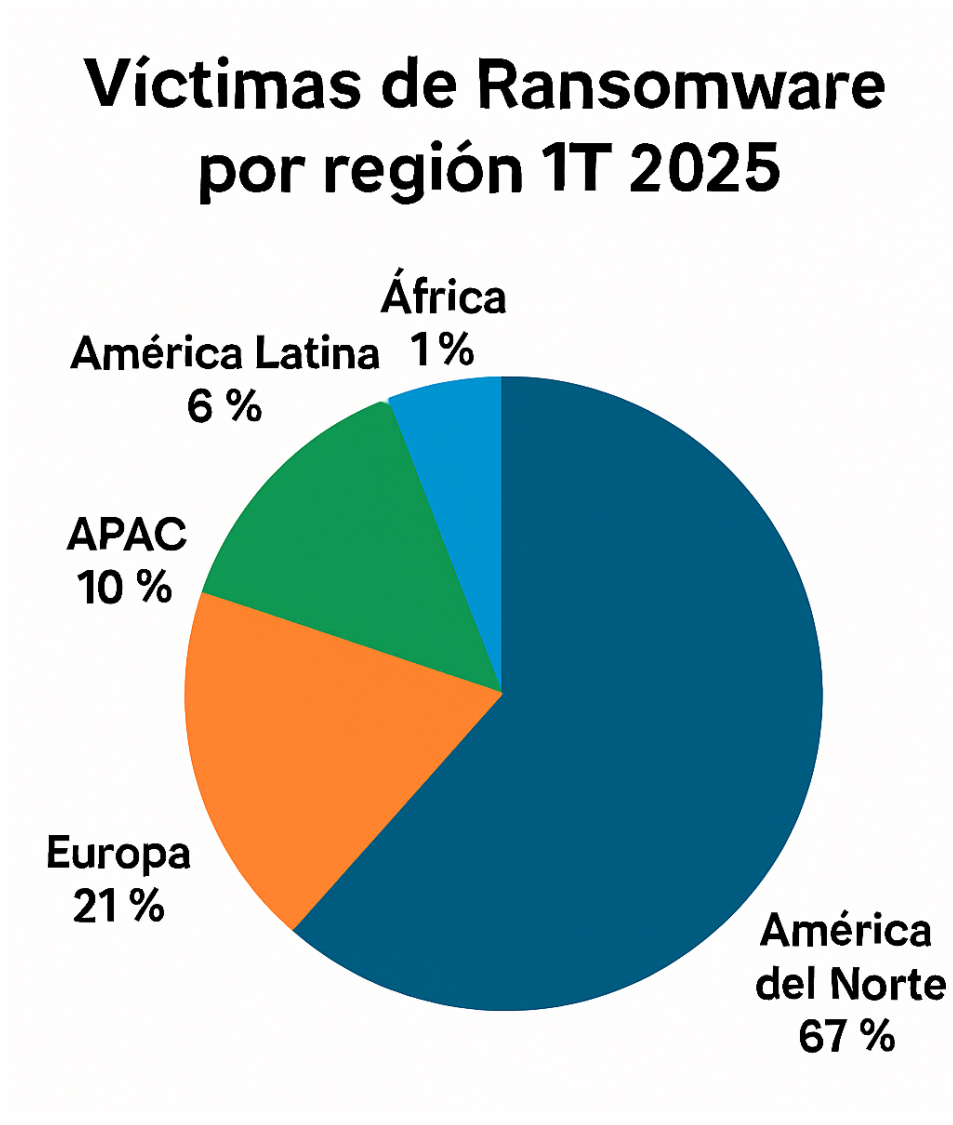

Figura 1. Víctimas de *ransomware* a nivel global. Fuente: Check Point, 2025

El espectro de las amenazas se ha expandido de forma considerable y sus operaciones se orientan hacia objetivos de alto impacto, como gobiernos, infraestructuras críticas y corporaciones de alto valor estratégico. El principal vector de ataque contra estos sectores

lo constituyen los grupos de Amenaza Persistente Avanzada (APT). Estas entidades suelen ser patrocinadas por Estados nación o por organizaciones criminales sofisticadas. Se especializan en ejecutar campañas de intrusión dirigidas y de larga duración, diseñadas para causar un impacto estratégico máximo. A diferencia de los cibercriminales convencionales, los grupos APT operan con un enfoque de largo plazo. Un ejemplo paradigmático de este tipo de grupos es Lazarus, una APT patrocinada por el estado de Corea del Norte. Atribuido a la Oficina General de Reconocimiento (OGR) (el principal órgano de inteligencia y operaciones especiales del país), el grupo es reconocido a nivel internacional por la magnitud y el impacto global de sus campañas de ciberataque. Se especializa en espionaje militar, operaciones clandestinas y actividades cibernéticas ofensivas.

En el año 2024 en el mundo se han producido ataques significativos de APT tales como:

• MirrorFace (Origen: Japón; alcance: Japón y Unión Europea), utiliza la herramienta HiddenFace. Emplea su línea de ataque a través de un correo electrónico de spear phishing donde MirrorFace incluye un enlace a un archivo ZIP. Dentro, con un LNK disfrazado como un documento de Word que una vez abierto, muestra un documento señuelo y ejecuta la versión 5.5.5 del *backdoor* ANEL.

• Gamaredon (Origen: Rusia, alcance: Ucrania, OTAN), emplea las herramientas: PteroGraphin, PteroPSDoor, PteroSig, su estrategia de ataque se basa en la realización de modificaciones a uno de sus *backdoors* escritos en PowerShell, conocido como PteroPSDoor. Estas mejoras incluyeron la incorporación de múltiples capas de ofuscación y el almacenamiento de componentes clave en el registro de Windows©, lo que aumentó de manera significativa su capacidad de operar de forma encubierta.

La «transformación digital» es el proceso mediante el cual las organizaciones utilizan tecnologías digitales para transformar de manera fundamental su forma de operar y de ofrecer valor a los clientes. Se trata de adoptar nuevas herramientas y métodos que puedan mejorar la eficiencia, optimizar la experiencia del cliente y crear nuevos modelos

de negocios. En conjunto con ello, se pueden mencionar desafíos, dentro de los cuales destaca la ciberseguridad. Las empresas, para lograr sortear esta amenaza, se deben apoyar en una plantilla detallada de informes de incidentes TI para garantizar una documentación exhaustiva y una respuesta rápida ante posibles brechas de seguridad.

Ataques ocurridos a sectores claves

• **Ataque de *ransomware* de la Corporación de Defensa Nacional: Fecha:** marzo de 2025, **Impacto:** 4,2 TB de datos confidenciales vulnerados. El grupo de Ransomware Interlock atacó a la National Defense Corporation (NDC) y a su filial AMTEC en marzo, exfiltrando 4,2 *terabytes* de datos que luego se filtraron en la *dark web*. **Aprendizaje clave:** Este incidente resalta la importancia de garantizar, el cumplimiento de las normas, de toda la cadena de suministro, incluidos proveedores internos y externos, además de los estándares de seguridad.

• **Ciberataque a WestJet: Fecha:** junio de 2025, **Impacto:** Interrupciones intermitentes en el sitio web y la aplicación móvil. En junio, la aerolínea canadiense WestJet reveló un incidente de ciberseguridad que interrumpió el acceso a su sitio web y aplicación móvil. **Aprendizaje clave:** Las organizaciones deben adoptar una defensa multicapa que incluya capacitación periódica en seguridad, especialmente en tácticas de ingeniería social. Capacitar a los empleados, especialmente a los equipos de TI y soporte, para que reconozcan los intentos de suplantación de identidad puede evitar que los atacantes se afiancen.

Tendencias de ataques cibernéticos

Los ataques cibernéticos se basan en el secuestro de datos. Hospitales, pequeñas y medianas empresas han sido las principales víctimas. A continuación, se relacionan una serie de ciberamenazas que, a nivel mundial se han evidenciado contra diferentes sectores y se prevé su aumento en los próximos años.

• **El ataque de denegación de servicio o DDoS:** consiste en provocar la caída de un servidor al sobrecargar su ancho de banda. Estas acciones fuerzan la interrupción de un sitio web. En el caso

del sistema financiero, los DDoS se utilizan para inundar con una gran cantidad de tráfico los servicios en línea de los bancos y de las plataformas de *trading*. De esa manera, el servidor colapsa y deja de funcionar.

• **Los troyanos bancarios:** otro *software* malicioso que en principio parecen inofensivos, sin embargo, no lo son y están tras los bancos. Los troyanos pueden instalarse en cualquier dispositivo por visitar un sitio web infectado, por descargar el anexo de un *email*, o incluso, por bajar una aplicación. Una vez este virus se instale en el celular detecta en qué momento se utilizan los servicios en línea de un banco y así capturar los datos personales y bancarios.

• **Vulnerabilidades *zero-day*:** los atacantes buscan fallos desconocidos en programas, dispositivos IoT o servicios *Cloud*. Estas brechas permiten el acceso a datos antes de que los desarrolladores tengan oportunidad de corregirlas.

• ***Malware* impulsado por IA:** la generación de código mediante inteligencia artificial (IA) ha democratizado y acelerado de manera significativa el desarrollo de *malware*. Los atacantes pueden generar, refinar y ofuscar variantes de *software* malicioso con una eficiencia sin precedentes, lo que facilita la evasión de mecanismos de detección basados en firmas y heurísticos.

• **Explotación de las API inseguras:** las interfaces de programación de aplicaciones (API) mal configuradas o sin la debida protección suponen una ventana para extraer datos, ya que muchas empresas se apoyan en servicios de terceros con controles de seguridad débiles.

• **Dispositivos IoT como puerta de entrada:** se pueden utilizar sensores, cámaras, periféricos y otros dispositivos conectados a la red corporativa como vectores de ataque para acceder a sistemas internos y extraer datos.

Blockchain, nube, teletrabajo

Blockchain

Las Tecnologías como *blockchain* aparecen como soluciones para mejorar la seguridad de datos y la confianza. Los ciberataques

dirigidos a la nube, *hardware* y sistemas de aprendizaje automático pueden crear brechas complejas, lo que dificulta a las empresas proteger sus datos. La tecnología *blockchain*, con sus libros digitales transparentes e inmutables, ofrece soluciones innovadoras para la ciberseguridad.

Las auditorías *blockchain* pueden abordar varios ataques y vulnerabilidades en contratos inteligentes. Para la gestión de respuesta a amenazas, *blockchain* permite la automatización y descentralización, lo que posibilita una rápida identificación, contención y mitigación de amenazas sin comprometer la seguridad de datos. Los contratos inteligentes pueden responder a ciberataques aislando las amenazas de forma automática y rápida. Desde la perspectiva de la ciberseguridad, la tecnología *blockchain* introduce avances transformadores: fortalece la resiliencia operativa, robustece los modelos de control de acceso y garantiza la integridad de los datos. Su arquitectura descentralizada mitiga el impacto de los ataques DDoS al eliminar puntos únicos de fallo, que dificulta de forma significativa la interrupción de servicios.

Al mismo tiempo, los principios de confianza cero (*Zero Trust*) se ven potenciados mediante mecanismos nativos de verificación de identidad y control de acceso granular, lo que reduce los riesgos de accesos no autorizados y amenazas internas. Además, el almacenamiento descentralizado de archivos (basado en la distribución de fragmentos cifrados a través de múltiples nodos) minimiza la exposición ante brechas de datos, mientras que la inherente inmutabilidad y replicación de la red aseguran la disponibilidad e integridad continua de la información, incluso durante incidentes de seguridad.

Nube

La seguridad en la nube es una disciplina de la ciberseguridad dedicada a asegurar los sistemas informáticos en la nube. Incluye mantener los datos privados y seguros a través de la infraestructura, las aplicaciones y las plataformas en línea. En su núcleo, la seguridad en la nube se compone de las siguientes categorías:

- Seguridad de los datos.
- Gestión de identidades y accesos (IAM).

- Gobernanza (políticas de prevención, detección y mitigación de amenazas).
- Planificación de la retención de datos (DR) y la continuidad del negocio (BC).
- Cumplimiento legal.

La seguridad en la nube es toda la tecnología, los protocolos y las buenas prácticas que protegen los entornos informáticos, las aplicaciones que se ejecutan y los datos almacenados en ella. La seguridad de los servicios comienza por comprender qué se está asegurando con exactitud, así como los aspectos del sistema que se deben administrar. El alcance total de la seguridad en la nube está diseñado para proteger lo siguiente:

- **Redes físicas:** enrutadores, energía eléctrica, cableado, controles de clima.
- **Almacenamiento de datos:** discos duros.
- **Servidores de datos:** *hardware* y *software* informáticos de la red central.
- **Plataformas de virtualización de equipos informáticos:** *software* de máquinas virtuales, máquinas anfitrionas y máquinas invitadas.
- **Sistemas operativos (OS):** *software* que soporta todas las funciones informáticas.
- **Middleware:** gestión de la interfaz de programación de aplicaciones (API).
- **Entornos de ejecución:** ejecución y mantenimiento de un programa en ejecución.
- **Datos:** toda la información almacenada, modificada y a la que se ha accedido.
- **Aplicaciones:** servicios tradicionales de *software* (correo electrónico, *software* de impuestos, paquetes de productividad.)
- ***Hardware* de usuario final:** ordenadores, dispositivos móviles, dispositivos de IoT.

Teletrabajo

El teletrabajo se consolidó como un pilar fundamental para la continuidad del negocio durante la pandemia, esto permitió

la operación de innumerables organizaciones. Sin embargo, esta modalidad ha expandido de forma crítica la superficie de ataque, transformando el acceso remoto de los empleados en un vector de amenaza frecuentemente explotado por actores maliciosos. Para enfrentar este panorama de riesgo, es imperativo la adopción inmediata de un conjunto de buenas prácticas de seguridad diseñadas para robustecer las defensas y minimizar la exposición a ciberataques, estas se describen a continuación:

- **Definir una política de seguridad para el teletrabajo:** siendo estrictamente necesario crear una política que defina los protocolos, orden y plan de crisis para cada uno de los escenarios en los que pueda verse involucrado el usuario.
- **Emplear modelo *Zero Trust*:** basado en las directrices, soluciones y políticas de seguridad para proteger el entorno de trabajo corporativo.
- **Gestionar la administración de dispositivos:** trabajar con un entorno de gestión basado en la Nube y aplicar las políticas adecuadas para su acceso.
- **Emplear una autenticación multifactorial (MFA):** contar con una base de identidad sólida que permite que los usuarios puedan acceder de forma segura a los recursos y aplicaciones allí donde quiera que se encuentren.
- **Trabajar en un entorno seguro:** acceder a recursos como servidores o conectarse los recursos de la empresa a través de VPN.
- **Implementar soluciones de seguridad integrada en la nube:** conectar todas las soluciones de identidad, aplicaciones y datos en un único ecosistema.
- **Realizar test de seguridad periódicos:** es necesario realizar periódicamente auditorías de seguridad.
- **Realizar copias de seguridad de manera automatizada:** realizar el *Backups* como medida de prevención y seguridad de la información para garantizar la continuidad ante cualquier incidencia que pueda afectar los sistemas.

Ciberseguridad en las redes 5G y 6G

La actual red 5G y su evolución hacia la 6G permiten una mayor velocidad de conexión, una menor latencia y nuevas posibilidades de interconexión masiva entre dispositivos IoT, así como el desarrollo de hologramas interactivos en tiempo real e internet sensorial, todo ello potenciado por la inteligencia artificial. Sin embargo, estas ventajas tecnológicas evidentes conllevan riesgos de seguridad paralelos. Por ello, todos los sectores de la sociedad deben prepararse para alcanzar la resiliencia necesaria ante las ciberamenazas y los ataques contra sus sistemas informáticos.

Dentro de los riesgos que implican el despliegue de esta se pueden citar:

1. Mayor superficie de ataque: la tecnología 6G promete la conectividad de más dispositivos, lo que se traduce en un aumento de posibles vulnerabilidades a ciberataques.

2. Compromiso de privacidad: esta tecnología incluye ajustes de localización y seguimiento en tiempo real de mayor precisión y esto compromete la seguridad de los usuarios.

3. Ataques más avanzados: las redes 6G necesitan de herramientas de IA para fortalecerse, pero, como ocurre con estas herramientas que se utilizan en la actualidad, pueden ser de ayuda para los ciberdelincuentes para llevar a cabo ataques más difíciles de detectar por los usuarios.

4. Vulnerabilidad de los dispositivos IoT: al incluir en su alcance a una gran cantidad de nuevos dispositivos IoT, si estos no cuentan con medidas de protección adecuadas, pueden ser objeto de ciberataques con más facilidad.

Estrategias generales ante los ciberataques

Los TELCOS y las empresas deben implementar medidas claves de ciberseguridad para proteger sus operaciones y la confianza de los clientes. A continuación, se relacionan algunas recomendaciones importantes a tener en cuenta:

- **Implementar sistemas de detección y respuesta avanzada:** detectar amenazas a tiempo es clave para prevenir daños mayores.

Los sistemas de detección y respuesta avanzada identifican patrones inusuales en el tráfico de red y responden de forma rápida a posibles ataques. Estas herramientas, basadas en IA y aprendizaje automático, son útiles para detectar amenazas desconocidas.

• **Adoptar la autenticación multifactor:** las contraseñas solas no bastan para proteger el acceso a sistemas sensibles. La autenticación multifactor agrega una capa extra de seguridad, lo que asegura que solo personal autorizado acceda a información crítica, algo esencial en el contexto de acceso remoto cada vez más común.

• **Capacitar a los empleados en prácticas de seguridad:** los ataques de ingeniería social, como el *phishing*, se aprovechan de la falta de conocimientos del equipo para acceder a sistemas y datos. Capacitar al personal en reconocer correos sospechosos y proteger contraseñas es clave para prevenir brechas de seguridad.

• **Actualizar y parchear con regularidad todos los sistemas y dispositivos:** uno de los errores más comunes es no mantener los sistemas actualizados. Las actualizaciones de *software* incluyen parches que corrigen vulnerabilidades conocidas. Es clave que las empresas de telecomunicaciones establezcan un calendario de mantenimiento regular para reducir el riesgo de ataques.

• **Implementar políticas de acceso y permisos estrictos:** limitar el acceso a sistemas y datos según el rol es una buena práctica de ciberseguridad. Las empresas deben asegurar que solo personal autorizado acceda a información confidencial y ajustar permisos según las funciones de cada empleado.

• **Realizar auditorías y evaluaciones de seguridad periódicas:** la ciberseguridad evoluciona y las amenazas cambian. Las auditorías y evaluaciones ayudan a identificar vulnerabilidades y ajustar medidas de seguridad, además de verificar la efectividad de las políticas y herramientas implementadas. La ciberseguridad es esencial en telecomunicaciones. En un entorno donde la conectividad y el intercambio de datos son clave, proteger la información y la infraestructura mantiene la confianza de los clientes y asegura el cumplimiento regulatorio. Medidas robustas y atención a nuevas amenazas fortalecen a cualquier empresa en el mercado.

Impacto de la IA en la ciberseguridad

La IA se ha consolidado como una herramienta poderosa que los defensores utilizan para anticipar, detectar, remediar y mitigar incidentes de seguridad con mayor eficacia. No obstante, los ciberdelincuentes también la emplean para automatizar la búsqueda de vulnerabilidades, generar *phishing* de mayor sofisticación y crear *malware*.

La convergencia de las tecnologías de *Fintech*; IoT con redes más rápidas, como la 5G, permitirá un volumen de tráfico muy superior al actual. Por ello, es necesario proyectar y emplear estas nuevas capacidades para defender los sistemas de ciberataques. A continuación, se relacionan los elementos clave que se deben considerar para el uso de la IA en este contexto.

Ataques más inteligentes y difíciles de detectar

- ***Phishing* hiperrealista:** con IA un atacante puede analizar perfiles sociales, correos electrónicos, imágenes y el lenguaje utilizado por una víctima para construir mensajes fraudulentos casi indistinguibles de los legítimos.
- **Automatización del reconocimiento de vulnerabilidades:** antes un ciberdelincuente examinaba de forma manual sistemas para encontrar vulnerabilidades. Ahora, con IA, puede analizar miles de componentes a gran escala y en tiempo récord para identificar vulnerabilidades y debilidades.

Recomendaciones

- **Herramientas de seguridad con IA integrada:** adoptar soluciones basadas en aprendizaje automático y análisis de comportamiento para detectar amenazas en tiempo real.
- **Supervisión humana indispensable:** los analistas con conocimientos sólidos y experiencia en ciberseguridad interpretan los resultados de la IA, identifican falsos positivos y negativos y toman decisiones informadas. La IA no es infalible y el factor humano sigue siendo imprescindible.
- **Trabajar con *partners* tecnológicos especializados:** muchas empresas optan por servicios gestionados de seguridad (MSSP) que utilizan IA para la monitorización y la respuesta a inciden-

tes aprovechando las capacidades y la experiencia de expertos especializados en ciberseguridad, sin necesidad de invertir en tecnología, formación e incorporación de un equipo interno.

Marco Normativo para la ciberseguridad

1) El Convenio de Budapest sobre la Ciberdelincuencia: este convenio es conocido formalmente como el Convenio sobre la Ciberdelincuencia del Consejo de Europa y se abrió a la firma el 23 de noviembre de 2001, como el primer tratado internacional que aborda los delitos cometidos a través de Internet y otras redes informáticas.

2) El Reglamento General de Protección de Datos (RGPD): este reglamento establece estándares estrictos para el procesamiento y almacenamiento de datos personales en la Unión Europea (UE).

3) Recomendación UIT-TX.1157 del 2015: describe las capacidades necesarias para el servicio de detección y respuesta al fraude de servicios basados en las TIC.

4) Ley de Portabilidad y Responsabilidad del Seguro de Salud (HIPAA): esta ley estadounidense entró en vigor en 1996 y tiene como objetivo proteger la privacidad y la seguridad de la información dentro del sector de la salud.

5) Estándar de Seguridad de Datos para la Industria de Tarjetas de pago (PCI DSS): se trata de un conjunto de normas y requisitos de seguridad diseñados para proteger la información de las tarjetas de pago y garantizar la seguridad de las transacciones con tarjeta de crédito y débito.

6) ISO 27001: se trata de una regla internacional que se centra en la seguridad de la información que proporciona un marco para establecer, implementar, mantener y mejorar un sistema de gestión de seguridad de la información (SGSI).

7) NIST Cybersecurity Framework: se trata de un conjunto de estándares, directrices y mejores prácticas diseñadas para ayudar a las organizaciones a gestionar y mejorar su postura de ciberseguridad. Se basa en tres principios fundamentales: gestión de riesgos, flexibilidad y escalabilidad, estructurado en cinco funciones:

- Identificar vulnerabilidades y brechas de seguridad de un sistema informático.
- Proteger los sistemas informáticos.
- Capacidad de detectar actividades maliciosas en un sistema informático.
- Saber cómo responder ante un incidente informático.
- Capacidad y recursos para recuperar la información ante un incidente informático.

8) Disposición Cubana sobre la Ciberseguridad: *Resolución* 105 Reglamento sobre el Modelo de Actuación Nacional para la Respuesta a Incidentes de Ciberseguridad, es la implementación de lo establecido en el Decreto 360/2019, tiene carácter preventivo y alcance a toda la sociedad, al implementar un sistema de trabajo entre las entidades especializadas en seguridad de las TIC para el cumplimiento de sus funciones en el intercambio seguro de información relativa a vulnerabilidades e incidentes de ciberseguridad.

Inversión en ciberseguridad

Se han abordado diferentes factores que impactan y constituyen riesgos, tendencias, proyecciones en el contexto de la ciberseguridad. Los avances tecnológicos requieren la inversión en ciberseguridad en las empresas de telecomunicaciones, que no se trata solo de proteger los activos, sino también de garantizar la resistencia y la sostenibilidad, al centrarse en áreas clave como la evaluación de riesgos, la gestión de proveedores, la planificación de la continuidad del negocio, la respuesta a incidentes y las pruebas continuas. La inversión debe enfocarse en robustecer los servicios de trabajo en la nube, *blokchain*, la computación cuántica y el empleo de herramientas de IA que permitan proteger a las infraestructuras de telecomunicaciones de amenazas o ataques.

Casos registrados durante este año de ciberataques a TELCOS

Telefónica

En enero de 2025, Telefónica, sufrió un ciberataque significativo que resultó en la filtración de aproximadamente 2.3 *gigabytes* de

datos internos. El ataque se originó a través de una sofisticada campaña de *phishing* dirigida a empleados de Telefónica. Los ciberdelincuentes emplearon técnicas de ingeniería social para engañar a los empleados y obtener sus credenciales de acceso. Una vez dentro del sistema, accedieron al sistema interno de *ticketing* de la empresa, conocido como Jira, y extrajeron una cantidad significativa de datos.

Consecuencias del ciberataque

1. **Filtración de datos internos y de clientes corporativos:** más de 236.000 registros de clientes empresariales y alrededor de 469.000 registros de *tickets* internos quedaron expuestos, lo que generó preocupaciones sobre privacidad y seguridad.
2. **Daño reputacional:** la exposición de información sensible afectó la imagen de Telefónica y disminuyó la confianza de sus clientes y socios estratégicos.
3. **Potenciales sanciones legales:** la filtración de datos puede derivar en sanciones económicas debido a incumplimientos de normativas como el Reglamento General de Protección de Datos (RGPD).
4. **Aumento de intentos de fraude:** los ciberdelincuentes pueden utilizar la información filtrada para realizar ataques dirigidos a los clientes afectados.

Respuesta

Una vez detectado el ataque, Telefónica activó su protocolo de respuesta a incidentes y trabajó con su equipo de ciberseguridad para mitigar el impacto y evitar nuevas filtraciones.

1) Se identificaron y bloquearon los accesos comprometidos. Esta acción permitió aislar las cuentas utilizadas por los atacantes y reforzar los controles de acceso.

2) Se estableció la obligatoriedad del restablecimiento de credenciales para todos los empleados. Esta medida se acompañó de la implementación de mecanismos más estrictos de autenticación multifactor.

Al mismo tiempo, la Empresa llevó a cabo un análisis forense del ataque con el objetivo de determinar qué vulnerabilidades fueron

explotadas y cómo mejorar su infraestructura de seguridad. Como resultado de esta investigación, Telefónica decidió fortalecer sus defensas internas, implementar soluciones de detección de amenazas en tiempo real y reforzar la segmentación de redes para dificultar la movilidad de posibles atacantes dentro de sus sistemas.

Lecciones aprendidas y estrategias de prevención

El incidente destaca la importancia de adoptar un enfoque proactivo en ciberseguridad. Algunas estrategias clave para prevenir ataques similares incluyen:

- **Autenticación multifactor (MFA):** evitar que el robo de credenciales sea suficiente para comprometer sistemas críticos.
- **Concienciación en ciberseguridad:** entrenar a los empleados para detectar intentos de *phishing* y aplicar buenas prácticas de seguridad.
- **Monitorización avanzada:** utilizar inteligencia artificial y análisis de comportamiento para detectar accesos inusuales.
- **Gestión de accesos con privilegios mínimos:** limitar los permisos a lo estrictamente necesario para cada usuario.
- **Pruebas de penetración periódicas:** evaluar la seguridad de los sistemas mediante simulaciones de ataques reales.

Claro Argentina

El 4 de julio del 2025, atacantes sustrajeron la base de datos de usuarios de Claro, una de las tres principales compañías de telecomunicaciones de Argentina. La información robada está actualmente a la venta en la *dark web*.

Cómo se realizó el ataque

El ataque tenía como objetivo la app y el portal de Claro, desde donde los usuarios gestionan sus servicios y realizan compras. Esto significa que los ciberdelincuentes no accedieron a las líneas de telefonía celular en sí, pero sí a cuentas asociadas a esos servicios, desde las cuales es posible realizar compras de equipos, solicitar servicios y gestionar información personal.

La filtración incluye correos electrónicos, nombres completos, direcciones, DNI, y códigos postales, datos sensibles que ya están circulando en la *deep web* y que también fueron subidos a sitios

como BiteBlob.com, una plataforma que permite compartir archivos de forma anónima. En esos foros, los datos se venden al mejor postor o incluso se publican de manera gratuita, lo que pone en riesgo la privacidad de miles de personas.

Acciones por parte de Claro

Se activaron los protocolos internos de análisis tras detectar que se mencionaba a la empresa en foros donde suelen ofrecerse bases de datos robadas.

Recomendaciones

Los expertos en general recomiendan una serie de medidas a los usuarios, entre ellas se encuentran:

- Cambiar contraseñas con regularidad y no repetirlas entre servicios.
- Activar la verificación en dos pasos en todas las cuentas.
- Evitar ingresar datos personales en sitios no oficiales o sospechosos.
- Estar atentos a posibles correos *phishing* y no descargar archivos desconocidos.
- Verificar con las empresas si sus cuentas han sido vulneradas y seguir las indicaciones oficiales.

Bouygues Telecom

La tercera mayor operadora de telecomunicaciones en Francia, con una base de alrededor de 27 millones de usuarios, confirmó el 4 de agosto haber sido víctima de un ciberataque. El incidente resultó en una brecha de datos que comprometió la información personal de un aproximado de 6.4 millones de sus clientes, lo que equivale a casi una cuarta parte de su base total de usuarios en el país.

Consecuencias

Los atacantes pudieron obtener información de contacto, detalles del contrato y su número de cuenta bancaria internacional (IBAN), entre los afectados se encuentran clientes particulares y empresas. No obstante, parece que los datos más confidenciales, como contraseñas y números de tarjetas de crédito, no se vieron afectados por esta violación.

Acciones

La operadora presentó una denuncia ante las autoridades judiciales. En ella, señaló que el autor podría enfrentarse a una pena de hasta cinco años de prisión y una multa de 150.000 euros. La TELCO no ha explicado el tipo de ataque, ni la vulnerabilidad atacada en su red.

Orange Francia

La multinacional francesa de telecomunicaciones Orange ha confirmado que fue víctima de un ciberataque el pasado 25 de julio de 2025.

Acciones

El ataque ha obligado a su equipo de seguridad, en colaboración con Orange Cyberdefense, a tomar medidas de contención inmediatas. El incidente, que afectó a uno de sus sistemas de información, ha tenido consecuencias en diversos servicios para empresas y consumidores, especialmente en Francia.

Medidas

La compañía presentó una denuncia oficial el 28 de julio de 2025, mientras las autoridades ya están involucradas en la investigación. No menciona indicios de que se haya producido una exfiltración de datos internos o de clientes. Apenas fue identificado el ataque, los equipos técnicos de Orange se movilizaron para aislar los servicios que podrían haberse comprometido. Esta acción preventiva tuvo como efecto colateral la interrupción de algunas plataformas de gestión para clientes empresariales y ciertos servicios para usuarios particulares.

Resultados y discusión

Los resultados obtenidos evidencian una creciente vulnerabilidad del sector de telecomunicaciones frente a amenazas cibernéticas cada vez más sofisticadas. En América Latina, se ha registrado un incremento del 108 % en los incidentes semanales, siendo México, Colombia y Brasil los países más afectados. Este dato revela una tendencia alarmante que exige una revisión profunda de las políticas de seguridad digital en infraestructuras críticas.

Entre las amenazas más frecuentes, el *ransomware* destaca como el vector de ataque predominante, que representa el 38 % de los incidentes globales. Variantes como LockBit y Akira han demostrado una capacidad de penetración significativa, lo que afecta tanto a grandes

corporaciones como a proveedores regionales. Además, se observa un aumento en la actividad de grupos de APT, como Lazarus y MirrorFace, que operan con objetivos estratégicos y respaldo estatal, lo que añade una dimensión geopolítica al problema.

Se muestra con la investigación el papel de tecnologías emergentes como la IA en la evolución de los ataques. La IA está siendo utilizada para generar *malware* más evasivo y automatizar campañas de *phishing* hiperrealistas. A su vez, la expansión de redes 5G y 6G, junto con el crecimiento exponencial de dispositivos IoT, ha ampliado la superficie de ataque, lo que genera nuevas vulnerabilidades en APIs y sistemas de autenticación.

El análisis de casos reales de ataques a Telefónica, Claro Argentina y Bougyes Telegom ilustra cómo las brechas de seguridad pueden comprometer la privacidad de millones de usuarios y dañar la reputación corporativa. Las respuestas implementadas por estas empresas, como la autenticación multifactor (MFA), la segmentación de redes y el análisis forense, constituyen ejemplos valiosos de buenas prácticas que pueden replicarse en otros contextos.

En cuanto al marco normativo, se destaca la importancia de instrumentos como el Convenio de Budapest, el Reglamento General de Protección de Datos (RGPD), la norma ISO 27001 y la Resolución 105 en Cuba. La cooperación internacional y la interoperabilidad entre gobiernos, operadores de telecomunicaciones y organismos reguladores se perfilan como elementos clave para una defensa efectiva y coordinada. Además, se evidencia el papel dual de la IA en este escenario. Por un lado, permite detectar amenazas en tiempo real y optimizar la respuesta ante incidentes. Por otro, plantea riesgos éticos y operativos al facilitar los ataques automatizados. En este contexto, la supervisión humana sigue siendo indispensable para interpretar los datos y tomar decisiones estratégicas.

En conjunto, los resultados confirman que la ciberseguridad ha dejado de ser una cuestión meramente técnica para convertirse en un componente esencial de la estrategia empresarial y nacional. La investigación aporta una visión integral del panorama actual y sugiere que futuras investigaciones deberían enfocarse en el impacto

económico de los ciberataques y en la evaluación de soluciones basadas en IA desde una perspectiva ética y de sostenibilidad.

Conclusiones

La ciberseguridad se consolida como un eje estratégico en el sector de telecomunicaciones, impulsada por la rápida adopción de tecnologías emergentes como *Fintech*, IoT, inteligencia artificial (IA), computación cuántica y redes 5G. Este ecosistema digital, cada vez más interconectado, ha ampliado de forma significativa la superficie de ataque, que ha dado lugar a amenazas más sofisticadas y persistentes.

En este contexto, la IA desempeña un papel doble: por un lado, potencia las capacidades defensivas mediante análisis predictivo y automatización de respuestas; por otro, habilita nuevos vectores de ataque más adaptativos y difíciles de detectar. Ante este panorama, se vuelve imperativo fortalecer las capacidades técnicas de los operadores, fomentar la interoperabilidad entre los distintos actores del ecosistema digital y promover una cultura organizacional resiliente, capaz de anticipar y mitigar riesgos. La colaboración interinstitucional, junto con el desarrollo profesional continuo, se posiciona como un componente esencial para enfrentar un entorno de amenazas dinámico, distribuido y en constante evolución.

Referencias bibliográficas

AECC Consultoras. (2025). *La transición al 6G pone a prueba la ciberseguridad global*. https://aecconsultoras.com/noticias-sectoriales/la-transicion-al-6g-pone-a-prueba-la-ciberseguridad-global/.

Benito, M. (2023). *Ciberdelincuencia: ¿qué es y cómo combatirla? Blog de los Estudios de Derecho y Ciencia Política*, Universidad Abierta de Cataluña. https://blogs.uoc.edu/edcp/ciberdelincuencia-que-es-y-como-combatirla/.

BitLife Media. (2025, agosto). *Qué sabemos del ciberataque a Orange: cronología, servicios afectados y respuesta de la compañía*. https://bitlifemedia.com/2025/08/que-sabemos-del-ciberataque-a-orange-cronologia-servicios-afectados-y-respuesta-de-la-compania/.

Check Point Software. (2025). *Informe global sobre ciberataques del primer trimestre de 2025* [Blog: Blog Check Point]. https://blog-checkpoint-com.translate.goog/research/q1-2025-global-cyber-attack-report-from-check-point-software-an-almost-50-surge-in-cyber-threats-worldwide-with-a-rise-of-126-in-ransomware-attacks/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc.

Djilp.org. (s.f.). *Marcos jurídicos internacionales sobre ciberseguridad y derecho de protección de datos*. Recuperado 18 de octubre de 2024, de https://djilp-org.translate.goog/international-legal-frameworks-on-cybersecurity-and-data-protection-law/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc.

El Siglo Web. (2025, julio 20). *Hackearon la base de datos de Claro y ya está a la venta en la dark web*. https://elsigloweb.com/2025/07/20/hackearon-la-base-de-datos-de-claro-y-ya-esta-a-la-venta-en-la-dark-web/.

Escudo Digital. (2025). *Un ciberataque a la tercera telco francesa expone los datos de 6 millones de clientes*. https://www.escudodigital.com/ciberseguridad/un-ciberataque-a-la-tercera-telco-francesa-expone-los-datos-de-6-millones-de-clientes.html.

Estrategias de Inversión. (2025). *Por qué invertir en ciberseguridad en 2025: claves, amenazas y oportunidades*. https://www.estrategiasdeinversion.com/fondos/por-que-invertir-en-ciberseguridad-en-2025-claves-n-818277.

Ministerio de Relaciones Exteriores de Cuba. (2021). *Cuba actualiza marco jurídico sobre telecomunicaciones y tipifica incidentes de ciberseguridad*. Misiones Cubaminrex. https://misiones.cubaminrex.cu/es/articulo/cuba-actualiza-marco-juridico-sobre-telecomunicaciones-y-tipifica-incidentes-de-0.

InfoSecurity México. (2025). *Una guía completa del concepto, tipos, amenazas y estrategias*. https://www.infosecuritymexico.com/es/ciberseguridad.html.

Kaspersky. (2025). *¿Qué es la seguridad en la nube?* Centro de Recursos. https://www.kaspersky.es/resource-center/definitions/what-is-cloud-security.

Piranirisk. (2025). *Ciberseguridad: qué es, cómo funciona y su importancia*. Academia Piranirisk. https://www.piranirisk.com/es/academia/especiales/ciberseguridad-que-es-como-funciona-y-su-importancia.

Revista Tono. (2023). "Comportamiento del fraude internacional y su impacto en operadores de telecomunicaciones", (17), 1-15. https://www.revistatono.etecsa.cu/tono/article/view/392/711

Telefónica Tech. (2025). "Tendencias en ciberseguridad para 2025." https://telefonicatech.com/blog/tendencias-en-ciberseguridad-para-2025

Winforsystems. (2025). Caso práctico: Ciberataque a Telefónica. https://winforsystems.com/caso-practico-ciberataque-a-telefonica/