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

tono

Scientific-Technical Journal of the Telecommunications Company of Cuba S.A.

ETECSA
EMPRESA DE TELECOMUNICACIONES DE CUBA S.A.

tono

Scientific-Technical Journal

Biannual Publication Vol.22 – No. 2 - 2025

RNPS: 0514 ISSN: 2224-6274

Tono, Scientific-Technical Journal of the Telecommunications Company of Cuba S.A.

The opinions expressed by the authors in the articles reflect their own points of view and do not necessarily coincide with the criteria of the editorial board.

The articles in this publication have been subjected to double-blind peer review.



Cover:
AI

Contact us

Scientific and Technical Surveillance and Information Division, ETECSA
Adress: Miramar Trade Center, 3rd Street between 76 y 78,
Beijing Building, 4th Floor, ofice 404. Playa, C.P.: 11300. Havana, Cuba.
Phone : (53) 7266-8453
Email: tono@etecsa.cu
Website: www.revistatonoetecsa.cu



Scientific Advisory Committee

Melissa Saltiel Delgado, M. Sc.
Mirta Julieta García García, M. Sc.
María del Pilar Caso Álvarez, M. Sc.
Alberto Javier García García, M. Sc.
Fidel Mirabal Puig, Eng.

Editorial Board

Editor-in-Chief
Grisel Ojeda Amador, M. Sc.
Executive editor
Inés María León Martínez, B. A.
Section editor
Alena Bastos Baños, B. A.
Proofreading
Mirta Ulloa Ferreiro, B. A.
Translation
Armando Camejo Hernández, B. A.
Frans Carlos Castellano Caballero, B. A.
Odalys Ojeda Fuentes, B. A.
Idalmis María Barbería Espinosa, B. A.
Alejandro Daniel Fadragas Robaina, B. A.
Design and Layout
Original design
Marcel Mazorra Martínez, B. A.
Layout
Isabel Canales Calvo, B. A.
Programming and Technical Assistance
Maritza de la C. Menéndez Figueredo, Eng.
Data review
Elisa María Alfaro Díaz, B. A.
Darian Díaz García, Eng.

Arbitrators

Caridad Anías Calderón, Ph. D. (CUJAE)
Alain Garófalo Hernández, Ph. D. (CUJAE)
Osvaldo Andrés Pérez García, Ph. D. (CENATAV)
Glauco Antonio Guillén Nieto, Ph. D. (ETECSA)
Armando Tito Bertot,M. Sc.(ETECSA)
Pablo Domínguez Vázquez, M. Sc. (ETECSA)
Dianelys Álvarez González, M. Sc. (ETECSA)
Maydel Vázquez Fernández, M. Sc. (ETECSA)
Karina Leonor Fernández Sánchez, M. Sc. (ETECSA)
María Carla Silveira Taboadela, M. Sc. (BIOCUBAFARMA)
Henry Raúl González Brito, M. Sc. (UCI)
Kevin Castro Rodríguez, Eng. (ETECSA)

EDITOR'S LETTER

On this occasion, a special edition of Tono Journal, which compiles the most significant contributions presented at the Cybersecurity in Telecommunications 2025 event, is published. This selection offers an updated overview of advances and challenges in this area, which is a priority for sovereign technological development.

Cybersecurity is positioned as a cross-cutting element in the digitization process of Cuban society. The studies in this compilation confirm its crucial role in protecting sensitive data, guaranteeing the secure operation of vital services, and maintaining the integrity of the national digital space. The researches provide coherent responses to contemporary threats that are becoming increasingly sophisticated.

In the national context, cybersecurity practices are distinguished by their systemic and integrative approach. The proposals presented at the event offer solutions that balance technological innovation and practical sustainability, paying special attention to the country's specific conditions. The achieved progress in areas such as operational security, data governance and specialized capacity-building is remarkable.

The event's findings confirm the need for an approach that combines technical improvements with civic skill development. It is noted that the effectiveness of digital protection strategies requires both advanced technological systems and conscious and prepared users.

This compilation proves the level of specialization and maturity achieved by this knowledge community in Cuba. It acknowledges the collective efforts of researchers and professionals whose work strengthens the foundation for secure and reliable national digital development.

Publishing Team
Tono Scientific-Technical Journal

TABLE OF CONTENTS

VOL.22 NO. 2
JULY / DECEMBER 2025

RESEARCH

Autonomic solution for cybersecurity integral management based on reinforcement learning 07
Ariel Baloira Reyes, Eng., Walter Baluja García, ScD, and Caridad E. Anías Calderón, ScD.

Zero-knowledge proof-based cryptographic authentication methods 19
Adiané Cueto Portuondo, Eng., Angel Alejandro Guerra Vilches, Eng. and Leanny Laura Duardo

Integration of security practices for DevSecOps pipelines in containerized environments 40
Mary Nelsa Bonne Cuza, Eng. and Meliza León Bencomo, Tech.

Deployment of honeypots for proactive detection of cyber threats 57
Heidy Rodríguez Malvarez, Eng., Elizabeth Molina Mena, Eng., Mónica Peña Casanova, Ph. D.

Proactive training against phishing emails using simulations with the GoPhish tool 66
Rubén Olabarrieta Rivera,Eng., Yuneisy Quintero Prieto, Eng., Raúl Sanchez Arañó, Eng.

Tool for monitoring and trace analysisin the Provincial Prosecutor's Office of Mayabeque 78
Laura Celín Sosa Martínez, ST.

Implementing artificial intelligence models for malware detection in critical telecommunications infrastructures 87
Elizabeth Molina Mena, Eng., Heidy Rodríguez Malvares, Eng. and Henry Raúl González Brito, M.Sc

SUMARIO

VOL.22 NO. 2
JULY / DECEMBER 2025

Integrated platform for digital signatures in the context of the company XETID 100
Humberto Martínez Rodríguez, PhD and Yosvel Dupeirón Porra, Eng.

CHRONOLOGY

The impact of new technologies on cybersecurity 112
Reniel Carvajal Alfonso, M.Sc. and Dianelys Álvarez González, M.Sc.

RESEARCH

Autonomic Solution for Comprehensive Cybersecurity Management Based on Reinforcement Learning

Ariel Baloira Reyes, Eng.[*], Walter Baluja García, ScD.[2], Caridad E. Anías Calderón, ScD.[3]

Received: 11/2025 | Accepted: 11/2025 | Published: 12/2025

Abstract

Cybersecurity has become a race against time. Cybercriminals improve their tools every day and launch multiple attacks in short intervals. This paper presents the functional components of an autonomous system for integrated cybersecurity management. Its main objectives are three: (1) to reduce detection, analysis, and response times to threats and vulnerabilities; (2) to automate cybersecurity management processes in Cuban entities; and (3) to maintain secure and stable infrastructure with minimal human intervention. Free and open-source software solutions were used to create the system. To provide the system with autonomy, two artificial intelligence models were deployed using Python and PyTorch. The development of this system constitutes a step forward for Cuba towards technological sovereignty in the area of cybersecurity.

Keywords: cybersecurity management, autonomic management, reinforcement learning, LLM, cybersecurity

[*] Engineer. Technological University of Havana “José Antonio Echeverría”, CUJAE. Calle 114 e/ Ciclovía y Rotonda, Marianao, La Habana, Cuba. ariel@tele.cujae.edu.cu

[2] ScD. Ministry of Higher Education. Calle 23 e/ F y G, Vedado, La Habana, Cuba. walter@mes.gob.cu

[3] ScD.Technological University of Havana “José Antonio Echeverría”, CUJAE. Calle 114 e/ Ciclovía y Rotonda, La Habana, Cuba.cacha@tesla.cujae.edu.cu

Introduction

The current cybersecurity has become a race against time between attackers and carriers of IT (Information Technology) infrastructures and systems. Threats have increased exponentially over the years, and their control represents a continuous wear on cybersecurity specialists (2024 SonicWall Cyber Threat Report, 2024; CrowdStrike, 2024; European Union Agency for Cybersecurity., 2024; Fabio Assolini, 2024). Cyber attackers have scaled up their operations through automation and the use of artificial intelligence (AI). This forces cybersecurity specialists to develop automated solutions to cope with them.

The research conducted by the Center for Strategic and International Studies (CSIS) in the United States shows that large organizations utilize approximately forty-seven different cybersecurity tools across their networks from an average of ten providers (Crumpler & Lewis, 2020). Managing this cluster of tools entails substantial monetary and human costs, especially to provide proactive and rapid threat response. As a solution to this problem, the integration of different technologies has been proposed, such as Security Information and Event Management (SIEM) systems, eXtended Detection and Response (XDR), AI, and cloud computing (European Union Agency for Cybersecurity., 2023; Uzoma *et al.*, 2023).

Timely detection and accurate identification of changes in infrastructure behavior patterns are crucial to ensuring the effective operation of cybersecurity protection mechanisms. The non-stationary nature of cyberspace poses a significant challenge for developing robust and up-to-date protection solutions. Therefore, the AI methods to be used need to dynamically adjust to the changing conditions of the environment so that defense systems respond effectively to threats. As a solution, the use of Reinforcement Learning algorithms is proposed, capable of generating models that imitate the human reasoning. These are adaptive, learn from the environment they are in, and select actions that maximize cumulative reward

based on prior experience (Adawadkar & Kulkarni, 2022; Alonso *et al.*, 2021; Hore *et al.*, 2022; Huang *et al.*, 2022; Mathew, 2021; Phan & Bauschert, 2022; Sewak *et al.*, 2022).

This paper presents a system based on free and open-source software tools to manage cybersecurity information, using RL techniques to respond autonomously to existing threats in the managed infrastructure. It allows cybersecurity specialists to oversee the system's decision-making and obtain a rapid and effective response to potential cyberattacks. The designed system seeks to achieve the following objectives: (1) to reduce detection, analysis, and response times to threats and vulnerabilities; (2) to automate cybersecurity management processes in Cuban entities; and (3) to maintain secure and stable infrastructure with minimal human intervention.

Materials and Methods

In this research, the scientific research methods of analysis-synthesis were used to study the components of SIEM and XDR solutions necessary to apply automated cybersecurity management in heterogeneous environments. Computational modeling and simulation methods were used to feign cyberattacks that the system must detect and mitigate automatically. Finally, empirical methods of observation, measurement, and experimentation were used during the validation of the proposed solution.

To build the system, the free and open-source software solutions Wazuh, OpenVAS, Suricata, and Snort were used. For the development of the AI solution, the Python programming language was used along with the PyTorch library as the main framework to develop the models. The proposed solution is in the development and optimization phase; however, there is a first version of the operating platform.

Results and Discussion

The system for automated cybersecurity management is composed of six components: (1) data collection, (2) monitoring and analysis, (3) detection and classification, (4) decision autonomy, (5) specialized supervision, and (6) storage. To guarantee stability and availability,

its design is based on a microservices architecture. This allows for the progression of components and enables simultaneous data collection by groups of collectors. The interrelation among components is shown in Figure 1.

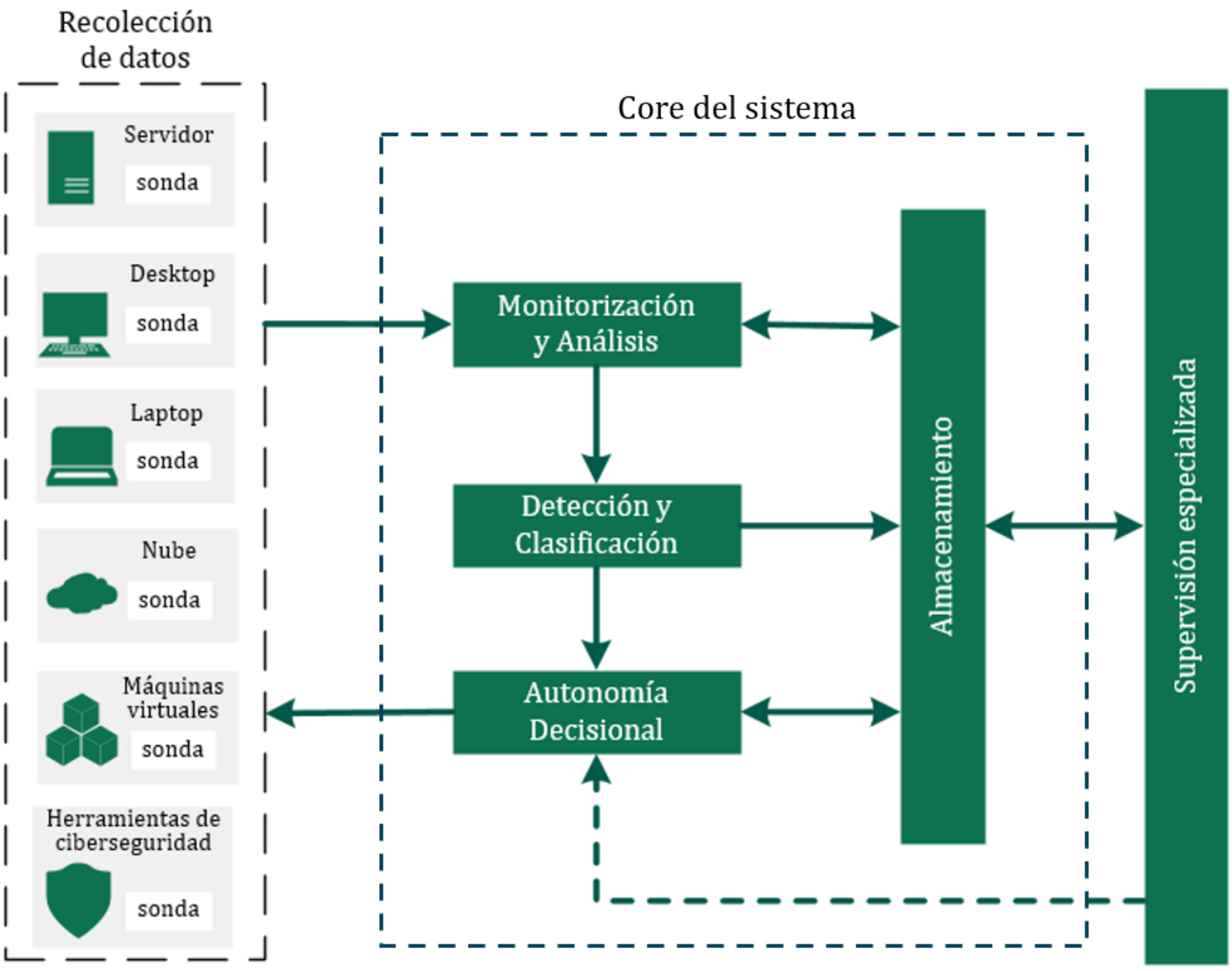


Figure 1. Internal components of the autonomous integrated cybersecurity management system

Data Collection

The main component is data collection, responsible for obtaining relevant information. It is composed of probes distributed across the network infrastructure and collection agents deployed on endpoints, called collectors. Its fundamental characteristic is the ability to gather heterogeneous data, including: operating system logs, network traffic, event logs from virtualization and fog/cloud computing instances, alerts from security systems such as Intrusion Detection Systems (IDS) or firewalls.

The collected data can be classified into two groups: general and specific. The general data consists of traffic logs, operating system logs, application logs, and security logs, while the specific data consists of vulnerability existence confirmations, threat data, and

collaborative intelligence data. It is important to emphasize that the specific type of data will vary depending on the cybersecurity tools used and the organization's needs; however, the general and specific data described above are considered essential to prevent and detect cybersecurity incidents. Specifically, the objective of the Data Collection component is to obtain raw cybersecurity data from various sources, in real-time or with minimal delay. These will be sent to the Monitoring and Analysis component so as to be processed and saved.

Monitoring and Analysis

To process and save the data collected by the collectors, it is necessary to have a common exchange point for the system. In this sense, the Monitoring and Analysis component provides standard communication interfaces with the Data Collection component. These are Syslog (System Logging Protocol), which allows the exchange of log messages in a standard format; and API REST (Application Programming Interface based on the Representational State Transfer architecture).

Multiple collectors may gather the same information and represent it differently. For this reason, one of its fundamental functions is the normalization of received information. This function is performed based on regular expressions (Zheng *et al.*, 2021) and the JSON (JavaScript Object Notation) format; a lightweight data interchange format easy for humans and machines to read and write (Banhara *et al.*, 2023). Its implementation allows generating a structured, easily understandable data structure, in addition to generalizing relevant information for analysis processes; based on this first step of raw data processing, logs can be enriched with associated information. For example, the IP address of a packet can reveal the corresponding country and service provider.

Another important function of the Monitoring and Analysis component is the correlation of data obtained from collectors to group information and avoid duplication. Given that network infrastructures can vary in size, this component must be capable of receiving and normalizing a high volume of information simultaneously.

Detection and Classification

It is responsible for detecting threats, anomalies and standard compliance violations and generating alerts when identifying any suspicious activity. This component is supported by multiple sources of threat intelligence and enriches alerts with contextual data to improve detection accuracy. This includes assigning events to the MITRE ATT&CK framework, detecting vulnerabilities with the Wazuh CTI service and aligning findings with regulatory standards such as PCI DSS, GDPR, HIPAA, the NIST 800-53 and CIS benchmarks (Figure 2). These capabilities provide practical information for threat detection, vulnerability detection, and compliance monitoring.

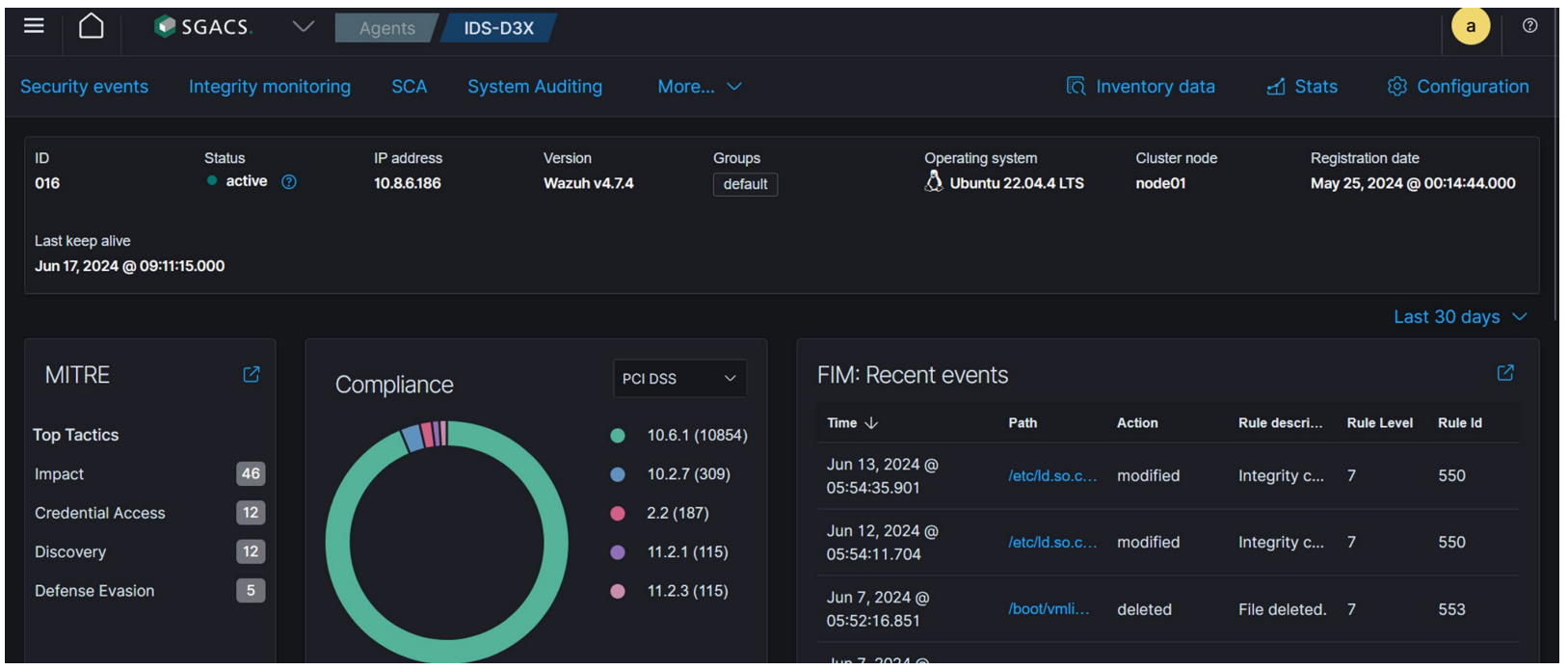


Figure 2. Window showing the operation of the Detection and Classification component

Using the Wazuh platform plug-ins, the Detection and Classification component is integrated with external platforms to streamline workflows. Examples include ticket systems like Jira or TheHive, as well as communication tools like Slack or Telegram.

Machine Learning (ML) models have a very important value. These are implemented as key pieces in threat detection and classification. For the implementation of this component, the algorithm kNN (k-nearest neighbors) was selected which is a supervised learning classifier that uses proximity to make classifications or predictions about the grouping of an individual data point.

Decision Autonomy

The brain of the system is represented by the Decision Autonomy component which is composed by two AI technologies: (1) Large Language Models (LLM) and (2) Deep Reinforcement Learning (DRL).

The LLM receives information from the Detection and Classification component, processes it, and organizes it into an observation matrix that serves as input for the DRL model. The latter decides what action to apply to reduce or mitigate risks. The integration of these models directly contributes to the system's autonomy.

The observation matrix represents the state of the managed infrastructure. Consequently, the LLM is trained and prepared to create an image of the current infrastructure situation. To achieve this, it relies on past events, data from the MITRE ATT&CK framework, data on present vulnerabilities, and the compliance status with international and national regulations. Given that the system is primarily intended for Cuban institutions, the LLM has been refined with national cybersecurity regulations. So far, experiments have been conducted with the Llama 3.2:3b model. Currently, a comparative study with other state-of-the-art models is being carried out to select the one that best fits the available resources.

On the other hand, studies conducted by Adawadkar and Kulkarni (2022); Asmat *et al*. (2025); Hore *et al*. (2022), Huang *et al*. (2022) and; Phan and Bauschert (2022) indicate that a DRL model possesses the ability to make strategic decisions based on both the current state and previous environmental conditions. To achieve this level of autonomy, a DDQN (Double Deep Q Network) type neural network was chosen, whose main advantage lies in its capacity for continuous improvement and its operation as a program that evolves similarly to human learning. In each iteration, the algorithm analyzes the actions performed based on the observed environment and adjusts its behavior with the aim of maximizing the reward obtained. Figure 3 shows the working diagram of this component.

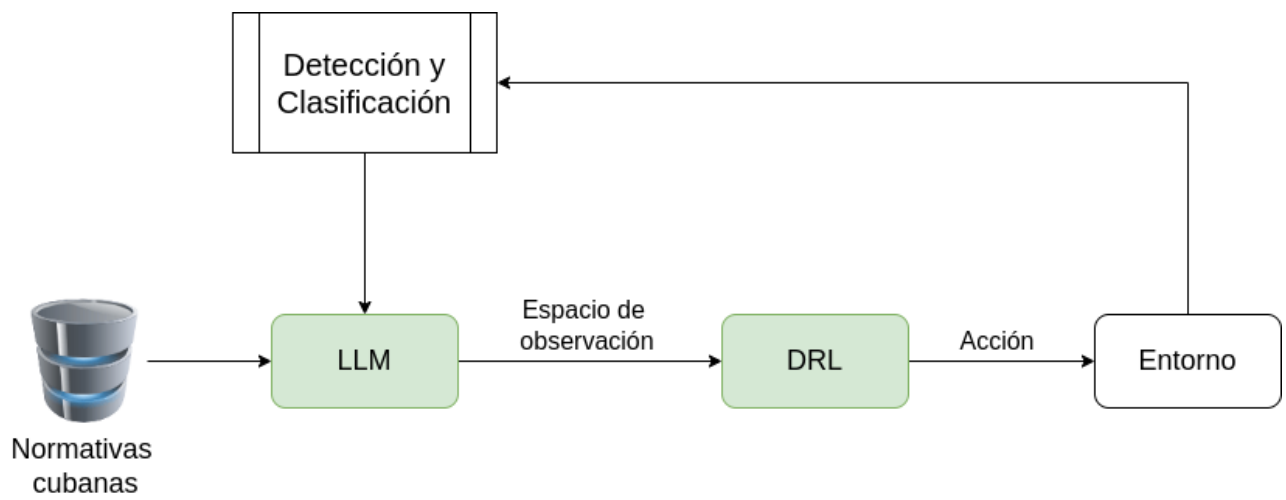


Figure 3. Diagram of the Decision Autonomy component

Storage

The Storage component manages all information generated by the above components, using JSON format for easy indexing. This feature is implemented using OpenSearch (https://opensearch.org/), an open-source search and observation platform. The primary function of the Storage component is data retention, regulatory compliance and forensic analysis. Being a distributed system deployed on the Docker technology, the component offers scalability and persistence. Data retention for long periods of time allows both specialists and researchers to access historical data and perform forensic analysis. In addition, it makes possible to analyze trends, identify recurring patterns and assess the effectiveness of implemented measures. At the same time, one of the primary cybersecurity requirements is guaranteed: the preservation of event data for analysis in situations that warrant it (*ISO/IEC 27001:2022, 2022*).

Specialized Supervision

Just as a teacher oversees the learning process of his/her students, it is necessary to implement control mechanisms that allow the system's decisions to be evaluated in real time and retrospectively. Despite the benefits of introducing AI in cybersecurity, the difficulty in understanding how these models act while operating autonomously is a risk that cannot be ignored by specialists or organizations (Arreche *et al.*, 2024; Atakishiyev *et al.*, 2024; Biswas *et al.*, 2024; Nwakanna et al., 2023; Rjoub *et al.*, 2023; Tsakalakis *et al.*, 2024). The Specialized Supervision component allows displaying the actions performed by the system, as well as accessing normalized and processed data, representing them clearly through diagrams, graphs, and tables (Figure 4).



Figure 4. Visualization of vulnerabilities present on a device

It is responsible for creating an intuitive graphical representation of the collection, storage, and processing of cybersecurity data. This representation can be leveraged by interested parties (such as security analysts, IT managers, and executives) to more clearly understand the organization's cybersecurity landscape and make informed decisions about risk management (Figure 5). This component makes it easier to interpret the elements that trigger a specific action in AI models, especially the RL one. Its function is essential to avoid overfitting and correct system deviations proactively.

›	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

Figure 5. Records of system actions and events

Conclusions

The development of an autonomous system for integrated cybersecurity management is another step towards national technological sovereignty. The integration of AI techniques allows reducing incident response times and correcting vulnerabilities before they are used. The six components that make up the system perform independent functions; however, when are integrated, they contribute significant capability to the developed solution. The Decision Autonomy component represents one of the main contributions to achieving autonomous cybersecurity, but it implies considerable risk if not managed properly. This control function is delegated to the Specialized Supervision component, which allows the visualization of all actions performed by the system and its controlled supervision. It is necessary to maintain the constant improvement of the language models used, in order to optimize resource usage. As a future line of work, it is essential to deepen into the study and design of explainability models aimed at interpreting the system's decisions and their implications.

References

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

RESEARCH

Zero-Knowledge Proof-Based Cryptography Authentication Methods

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

Received: 11/2025 | Accepted: 11/2025 | Published:12/2025

Abstract

This article examines the evolution of authentication methods in a context where security and privacy are increasingly crucial. The need for systems that guarantee the confidentiality and integrity of information is highlighted, especially in a vulnerable digital environment. The study focuses on exploring different authentication categories, from traditional methods such as passwords and physical tokens to more advanced approaches such as Multi-Factor Authentication (MFA) and the use of digital certificates. However, the main focus lies in the analysis of Zero-Knowledge Proof (ZKP) as an innovative solution to mitigate the vulnerabilities inherent in traditional methods. The methodology used includes a comprehensive literature review and a detailed analysis of ZKP protocols applied to authentication, with special emphasis on their

[*] Cybersecurity Department, Cybersecurity Faculty, University of Computer Sciences. adianecp855@gmail.com

[2] Technological Infrastructures Department, Cybersecurity Faculty, University of Computer Sciences. angelagv@uci.cu

[3] Technological Infrastructures Department, Cybersecurity Faculty, University of Computer Sciences. leannyldp@uci.cu

implementation in cloud computing environments. The results emphasize the effectiveness of ZKP when verifying user identities without compromising sensitive information and provide significant advantages compared to other methods.

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

Introduction

An average citizen uses technologies to access all kinds of services (email, bank accounts, media networks, mobile devices, communications services, among others). These services must provide security and reliability, for which some type of information that uniquely identifies the user is required (Rodríguez Valdés *et al.*, 2018).

Authentication is a process that verifies the identity of a user attempting to access specific information. Authentication systems are routinely used to protect sensitive data and prevent unauthorized access to systems and networks (Marmolejo-Corona *et al.*, 2023). This process is fundamental in order to guarantee the confidentiality, integrity, and availability of resources in a digital environment, as effective authentication minimizes the risk of unauthorized access and protects information against theft, alteration, or misuse. With the increase in cybercrime and the exposure of personal information, the need for robust and secure authentication methods has become crucially important. Authentication techniques have evolved from traditional passwords to Zero Knowledge Proof (ZKP) to deal with security challenges in an increasingly digitalized world.

This work explores various authentication methods, from the most traditional to the most innovative, with a special focus on zero-knowledge cryptography. The main problem addressed by this research is the vulnerability of traditional authentication methods (such as passwords and physical tokens) against sophisticated attacks and the exposure of sensitive information. This paper is divided into several sections. First, a comprehensive literature review on authentication methods is carried out, highlighting the limitations

and vulnerabilities of traditional approaches. Secondly, it delves into ZKPs, explaining their theoretical foundations and their application in authentication. Different ZKP protocols (such as Schnorr's, Feige-Fiat-Shamir's, and Guillou-Quisquater's) are analyzed, and their security and efficiency are discussed.

The ultimate goal is to provide a deep understanding of how ZKP can be used to secure user identity without compromising privacy or information security. This innovative approach not only mitigates the vulnerabilities of traditional methods but also opens new possibilities for secure authentication in an increasingly complex digital environment exposed to threats.

Materials and Methods

In order to develop this study, an exhaustive search was conducted in various recognized academic databases such as Google Scholar, Scopus, IEEE Xplore and ACM Digital Library

With the aim of ensuring the inclusion of relevant and high-quality studies, specific inclusion criteria were established.

- Publications in peer-reviewed journals or high-impact conferences: studies published in scientific journals with a rigorous peer-review process were prioritized, as well as those presented at internationally recognized conferences in the fields of cryptography, cloud security, and data privacy.
- Focus on authentication methods based on ZKP: studies addressing the development, implementation, and evaluation of authentication techniques employing ZKP were selected.
- Applications in cloud security and blockchain: works researching the application of ZKP in the context of cloud security and blockchain technologies were prioritized; use cases such as authentication in distributed systems, privacy protection in transactions, and data integrity are highlighted.

The methodology employed in this research was divided into two complementary approaches. First, a literature review was conducted, exploring the existing literature on authentication methods based on ZKP in the context of cloud security and blockchain. This review

provided an overview of the current state of research in this field and identified potential areas for improvement or innovation.

Secondly, an analytical-synthetic approach was applied, involving the detailed analysis of the theoretical principles underlying ZKPs and their application in cryptographic authentication. Through a literature review, key concepts were explored in depth and the acquired knowledge was synthesized to identify potential areas for improvement or new research directions. This approach combined theoretical understanding with the identification of practical applications, providing a solid basis for progress in the field of zero-knowledge cryptography-based authentication.

There are various methods to authenticate users that vary in security and complexity. These can be grouped into three broad categories:

1. **Knowledge-Based Authentication**

   Alphanumeric Passwords

   Alphanumeric passwords are among the most traditional and widely used methods within knowledge-based authentication. They are a combination of characters that use the ASCII code to generate a secret key of variable length. They are easy to implement and very common in all types of systems. A strong alphanumeric password should be random and long, therefore difficult to remember. Due to the nature of mnemonic processes when users have to establish or remember many passwords, which they use in different systems, they tend to use similar keys or phrases that are easy to remember (Rodríguez Valdés *et al.*, 2018). This method, although the most common, is extremely vulnerable to various types of attacks, including:

   - Phishing: where hackers trick users into revealing their passwords through fake websites or malicious emails.
   - Dictionary attacks: which use predefined lists of common words and alphanumeric combinations to try to guess the password.
   - Brute force attacks: which involve trying all possible combinations until the correct password is found.

One-Time Password (OTP)

In contrast to static passwords, another common method within the same knowledge-based authentication group is the One-Time Password (OTP). This system ensures the use of passwords with an OTP system, meaning that the user has a valid password for a specific time and purpose. This solution is generally used for the initial authentication process for external accesses, which are carried out via Virtual Private Network (VPN) (Mendoza Arteaga *et al.*, 2020). Although OTP systems offer an additional layer of security by generating temporary passwords that expire after a single use, they also present vulnerabilities that hackers can exploit. For example, if a hacker manages to intercept the communication channel (such as SMS messages or emails) through which the OTP is sent, they could use the code before it expires and access the target account or system. Furthermore, the mobile devices used to receive OTP can be compromised by malware, allowing hackers to capture the codes in real time.

Another significant vulnerability is targeted phishing, where a hacker tricks the user into revealing their OTP on a fake website or through a malicious application. In this way, hackers can authenticate as if they were the legitimate user. These risks highlight the need to implement additional security measures.

Graphical Authentication

Graphical Authentication Systems are used both for user authentication in a system and for key generation for use in cryptographic algorithms (Shendage, S. *et al.*, 2014). Graphical passwords can be made up by a combination of photos, images, or iconography. The characteristics of the image produce a much larger key space. The efficiency of these systems is based on the great capacity of humans to remember patterns in images, instead of memorizing sets of characters of great length and complexity (Rodríguez Valdés *et al.*, 2018).

In general, these systems have proven advantages over traditional text input methods and maintain considerable key spaces. However, graphical authentication also has some significant disadvantages.

One of its main drawbacks is that its implementation can be more complex and costly than traditional methods, as it requires more advanced graphical interfaces and efficient image storage and retrieval systems. Furthermore, they depend on devices with modern graphical capabilities, which can limit their use on older or resource-constrained devices. The use of personal or identifiable images can raise privacy concerns, and although key spaces are larger, brute force attacks remain a threat. The user interface may be less intuitive for some users, and compatibility between different systems and platforms can be an issue.

2. Possession-Based Authentication

Physical Tokens

In addition to what the user knows, possession-based authentication relies on what the user possesses, such as physical tokens. A token is a storage device that contains a secret key used in the authentication process (Mendoza Arteaga *et al.*, 2020). Tokens are issued and certified by a trusted entity, such as a certification authority, which validates the authenticity and integrity of the information contained in the token, adding an additional layer of security.

Generally, these tokens are as smart cards or magnetic cards, but they can be seen in many other forms such as digital devices or USB drives. These cards contain a microchip capable of storing information and performing cryptographic functions, allowing their use in a wide range of applications, from e-commerce to physical and logical access control (Mendoza Arteaga et al., 2020). In many countries, smart cards have become an integral part of identity documents, facilitating not only personal identification but also secure authentication in online and in-person services.

In addition to the risks of loss or theft, tokens (like any physical device) can deteriorate over time due to continuous use. This deterioration affects the token's functionality, making it less reliable, or even unusable in some cases.

3. Inheritance-Based Authentication

Biometrics

Within authentication methods, biometrics is an approach based on who the user is. This method consists of authenticating an individual through their physical characteristics: facial features, fingerprints, voice structure and pattern, iris pattern, and other unique characteristics of each living being. Biometric recognition involves storing, measuring, and comparing an individual's distinctive features. (Mendoza Arteaga *et al.*, 2020).

Authentication schemes based on biometric information are alternatives to alphanumeric passwords. Their implementation requires, in all cases, the use of specialized hardware elements. This requirement raises the systems' costs and complexity, which can be a barrier to their use in resource-constrained environments. Additionally, the accuracy and reliability of these devices can be affected by external factors, such as hardware quality, environmental conditions, or the user's physical state. Biometrics can be conceptually difficult to forge, but easy to spoof. This does not mean they can be copied or duplicated as easily as a password, but the collection of biometric data (such as fingerprints or facial images) can be done without the user's knowledge in high-tech situations. The performance of these systems can also be compromised by circumstances such as the user's health, stress, and other factors that can make the process lengthy and uncomfortable (Rodríguez Valdés *et al.*, 2018).

Multi-Factor Authentication (MFA)

A single authentication factor is not secure, so currently Multi-Factor Authentication (MFA) is the trend for verifying that users are legitimate (Mendoza Arteaga et al., 2020). This security process allows the users to corroborate their identity by confirming two different authentication factors (static key and dynamic key) (Espinoza Tinoco *et al.*, 2022). For example, a system may require a password (something the user knows) and a code sent to her/his phone (something the user has). MFA significantly increases security by requiring multiple levels of authentication.

This type of authentication has vulnerabilities related to the way it is implemented. If a website uses an SMS-based two-factor authentication method, hackers can clone the user's SIM card and redirect text messages to their own device. Additionally, if an authentication app is used, the user's device can be compromised by malware or viruses, allowing hackers to access the security codes generated by the app (Marmolejo-Corona *et al.*, 2023).

The implementation of MFA can create a conflict between convenience and security. For many users, the need to complete additional steps to authenticate can be perceived as inconvenient and cumbersome. This can lead to resistance to their use, or even deactivation of these measures if given the option. Moreover, not all online services are compatible with MFA, forcing users to choose between security and convenience. This incompatibility can also create technical difficulties in configuring or using MFA, especially on outdated devices and applications.

Another significant challenge of MFA is its reliance on third-party providers to deliver authentication codes or validate biometric data. This subordination introduces an additional point of failure in the authentication process. If the external service is compromised, the user's authentication may also be at risk, exposing the system to potential security breaches.

Digital Certificates

Among other more recent methods is the digital certificate. This consists of an electronic document that verifies the identity of a person or organization on the Internet and uses cryptographic technology to protect the privacy and security of online communications. A certification authority issues it and is responsible for verifying the identity of the certificate registrant and ensuring the authenticity of the public key associated with the certificate. The certificate enables the encryption and decryption of data, the digital signing of documents, and the authentication of users in online transactions and services (Marmolejo-Corona *et al.*, 2023).

Although essential for online security, digital certificates can be vulnerable if the associated private key is compromised. In such a

case, a hacker can impersonate the user's identity, leading to serious security risks. Furthermore, they have an expiration date and if not renewed in time, critical services may be interrupted. Trust in certification authorities (CA) is essential; however, if a CA is breached, the entire certification system is affected. Finally, complying with regulatory rules and standards requires constant management and can be costly for organizations.

Authentication through the use of Zero-Knowledge Proofs

Despite advances in authentication methods (such as multi-factor authentication and the use of digital certificates), challenges related to security and privacy persist. Traditional methods, although effective in many cases, remain vulnerable to sophisticated attacks and the exposure of sensitive information. This is where zero-knowledge proofs emerge as an innovative solution that addresses these issues in a unique way.

ZKPs allow one party (prover) to demonstrate to another (verifier) that they possess certain knowledge or have performed an operation without revealing any additional information about that knowledge or the process itself. This approach resolves one of the key limitations of traditional authentication methods: the need to share sensitive information to validate the user's identity.

Authentication through ZKPs falls mainly within knowledge-based methods, since the user proves knowledge of certain information without revealing what that information is. In this context, knowledge is what enables authentication, but unlike traditional methods (such as passwords), the ZKP protocol guarantees that the knowledge is not shared with the authentication system, adding a crucial layer of security (Fenzi, 2019).

Cloud Authentication

Cloud computing offers quick and economical access to computing resources such as storage and applications, making it attractive for small and medium-sized enterprises. However, this model has serious security challenges such as privacy issues, lack of transparency, data leaks, and identity theft. The connected and dynamic nature

of the cloud increases vulnerability to attacks, as traditional defenses are insufficient to protect this environment.

Cloud services require users to store sensitive information, exposing data to potential unauthorized access, especially when weak authentication methods are used. Furthermore, the multiple copies of data and services in the cloud create a complex environment that can result in management and security problems, increasing the risk of attacks such as session hijacking, which involves obtaining credentials during login (Bermúdez, M. *et al.*, 2020).

Given this scenario, a robust cloud security approach that outdoes traditional practices is required. This is where ZKPs present themselves as an effective solution. In this context, ZKPs allow verifying user identities without revealing sensitive information, mitigating the risks associated with data exposure. Their implementation provides an additional layer of security by preventing hackers from accessing critical information, even if they manage to intercept communications. This not only protects the integrity and confidentiality of data but also effectively addresses the main security concerns associated with cloud computing.

Zero-Knowledge Proofs (ZKPs)

The ZKP cryptographic protocol allows sharing and verifying information without exposing sensitive data. Essentially, this protocol outlines a procedure in which one party (the prover) claims that something is true and another party (the verifier) checks that it is true without knowing all the details. Therefore, it is a system that proves knowledge of something that has happened without the proof requiring revealing all the data (Guillou & Quisquater, 1988).

The above is achieved by generating a proof by the prover that meets a specific set of criteria, which the verifier can use to confirm the claim without learning anything more about the statement. Consequently, ZKPs allow minimizing and limiting access to data in distributed contexts, such as general Internet services or cloud computing (Vilchez Moya *et al.*, 2023).

Zero-knowledge cryptography is based on the concept of a "convincing deceiver," where one entity, the verifier, can be convinced that another entity, the prover, possesses certain information without learning anything more about it. This is achieved through interaction between the verifier and the prover, in which a series of challenges and responses are carried out that allow verifying the validity of the claim without revealing additional details (Hernández, D. *et al.*, 2020).

Here is an example: Alice and Bob are facing a particular challenge. Bob is colorblind and cannot distinguish between red and green. Alice has two identical circles that differ only in color: one is red and the other is green. Alice's goal is to prove to Bob that the circles are different without revealing any additional information. In this scenario, Alice is the prover and Bob is the verifier.

1. Alice shows the two circles to Bob and explains that they are different, but Bob cannot differentiate them.
2. To start the exercise, Alice asks Bob to place both rings behind his back. She then instructs him to show one of the rings and hide it again.
3. Next, Alice instructs Bob that he has the option to show the same circle he initially showed again or switch it for the other circle.
4. Each time Bob shows a new circle, Alice instructs him whether he switched the circle or not.
5. Alice correctly guesses whether Bob switched the circle or not in each successive turn.
6. As Alice guesses correctly, it becomes increasingly probable that the circles are different.
7. Through this interaction and Alice's consecutive successes, Bob becomes convinced that the circles are different without Alice having to explicitly reveal why.

A zero-knowledge proof must satisfy three properties:

- Completeness: if the statement is valid, then the honest verifier (Bob) will be convinced of this fact by an honest prover.

Consequently, Bob will accept Alice's identification with probability one.

- Solidity: if the statement is invalid, no cheating prover can convince the honest verifier that it is true. Consequently, the verifier (Bob) will reject the proof with a probability of at least one-half.
- Zero-knowledge: if the statement is true, no cheating verifier can discover anything new apart from this information (Ezziri & Khadir, 2020).

There are various zero-knowledge protocols that share the principle of proving that one knows something without having to specify exactly what that knowledge is. In the interactive zero-knowledge protocol, the verifier and the prover interact with each other multiple times before the verifier can be sure that the prover knows the required information. The protocol is based on the assumption that if a user knows the information, this user can answer certain questions or perform certain operations correctly. An imposter would not be able to do so. Consequently, interactive ZKPs have applications in various areas, such as authentication, identity verification, privacy protection, and security in distributed systems (Henry, R., 2014).

In contrast, in the Non-Interactive Zero-Knowledge (NIZK) protocol, the prover sends a single proof to the verifier, without any additional interaction being necessary.

In this type of protocol (such as the Schnorr's proof), the party that knows the secret interacts with another party (the verifier) to prove that they know the secret without revealing it and without the need for interaction between them. Examples of systems based on NIZK include cryptocurrencies such as Zcash and Monero, as well as two-factor authentication solutions based on public-key cryptography such as Universal Second-Factor (U2F) authentication.

Another type of zero-knowledge protocol applies graph theory to prove knowledge of certain relationships between elements of a graph. This protocol is based on a challenge-response between the two parties. In the first phase, the party wishing to prove that it possesses the information (proponent) constructs a graph from said information and

sends it to the other party (verifier). In the second phase, the verifier chooses a node from the graph and sends a question to the proponent asking to provide the information related to that node. In the third phase, the proponent answers the question without revealing additional information about the graph. The process is repeated several times on different nodes of the graph in each iteration. Ultimately, if the proponent can correctly answer all the verifier's questions, it is considered to have proved that he/she know the information without revealing it.

Finally, it is worth mentioning the Elliptic Curve-Based Zero-Knowledge protocol, a very secure encryption system that uses complex mathematics to protect and decrypt information. In this system, a distinction is also made among the sender, the verifier, and the proof. The latter is the process that allows verifying the authenticity of the information without revealing it, and to carry it out, the sender and verifier select an elliptic curve and a starting point on that curve. The sender then performs a series of mathematical calculations using the information to be verified and sends the results of these calculations to the verifier, who uses these results to determine whether the information is authentic or not. This protocol is very fast and efficient, making it ideal for use in online systems and other high-speed environments (Wanden-Berghe, C.A., 2013).

Protocols Used in Authentication

Schnorr's Protocol

The Schnorr's authentication and signature scheme derives its security from the difficulty of computing discrete logarithms. To generate a key pair, two prime numbers (p and q) are first chosen, such that q is a prime factor of $p-1$. Then, a number not equal to 1 is chosen, such that $a^q \equiv 1 \ mod \ p$. All these numbers can be common to a group of users and can be freely published. To generate a specific key pair (public key/private key), a random number less than q is chosen. This is the private key s. Then, $v \equiv a^{-s} \ mod \ p$ is calculated.

This is the public key (Schneier, 1996). Schnorr's Authentication Protocol (See Figure 1):

1. Alice chooses a random number *r* less than q and calculates $x = a^r \ mod \ p$. This is the preprocessing stage and can be done long before Bob is present.
2. Alice sends *x* to Bob.
3. Bob sends Alice a random number *e* between 0 and $2^t - 1$.
4. Alice calculates $y = (r + s \cdot e) \ mod \ q$ and sends it to Bob.
5. Bob verifies that $x = a^y \cdot v^e \ mod \ p$.

This is because $x = a^{r+s \cdot e} \cdot (a^{-s})^e = a^r \ mod \ p$.

Security is based on the parameter *t*. The difficulty of breaking the algorithm is 2^t approximately. Schnorr recommended that *p* be around 512 bits, q around 140 bits, and t be 72 bits.

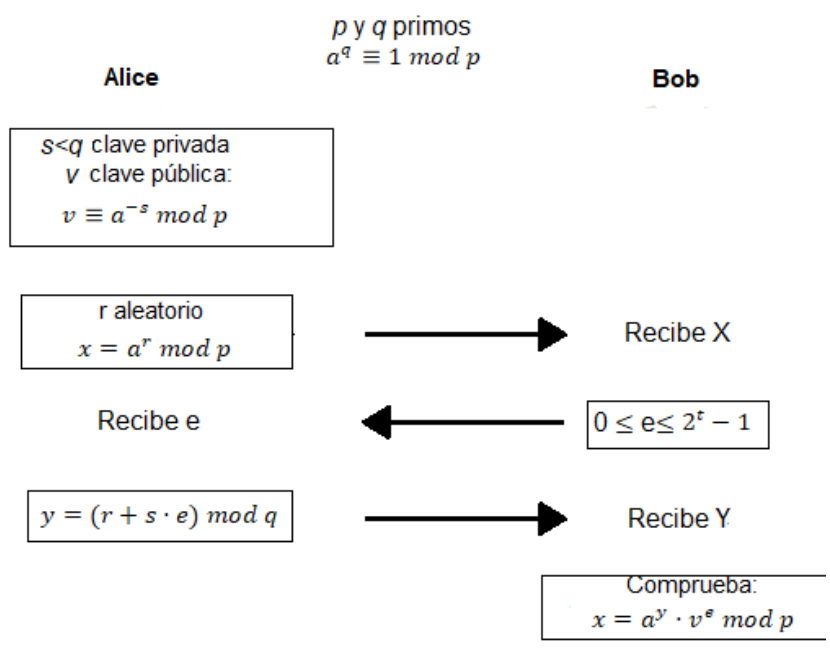


Figure 1. Schnorr Protocol Round

Since calculating the discrete logarithm *s* from $v \equiv a^{-s} \ mod \ p$ is computationally difficult, the protocol ensures that a hacker cannot derive the private key from the public key.

Fiege-Fiat-Shamir's Identification Protocol

The protocol created by Uriel Fiege, Amos Fiat and Adi Shamir in 1988 is one of the most common implementations of a ZKP. The Fiat and Shamir protocol is based on the difficulty of the square root. The applicant proves knowledge of a square root modulus a large number *n* (Hasan, J., 2019).

The Fiege-Fiat-Shamir Identification Scheme includes two entities (the prover and the verifier), according to the principle of all

ZKPs. The prover has a secret token and is looking for authentication, so he/she must prove to another entity (the verifier) to authenticate him/her based on the prover's secret token through a sequence of problems, without the verifier knowing the prover's secret token.

Single configuration

A trust center publishes a modulus $N = p \cdot q$ which is the product of two prime numbers. Each potential applicant (proponent) selects a secret s that is coprime with N and publishes $v = s^2 \ mod \ N$ as their public key (Figure 2).

1. Public: the modulus N and $v = s^2 \ mod \ N$.
2. Alice selects r and sends to Bob $x = r^2 \ mod \ N$.
3. Bob chooses e ∈ {0,1}.
4. Alice sends $y = r \cdot s^e \ mod \ N$.
5. Bob must verify: $y^2 = x \cdot v^e \ mod \ N$.

This is because $y^2 = r^2 \cdot s^{2e} = r^2 \cdot (s^2)^e = \ x \cdot v^e \ mod \ N$.

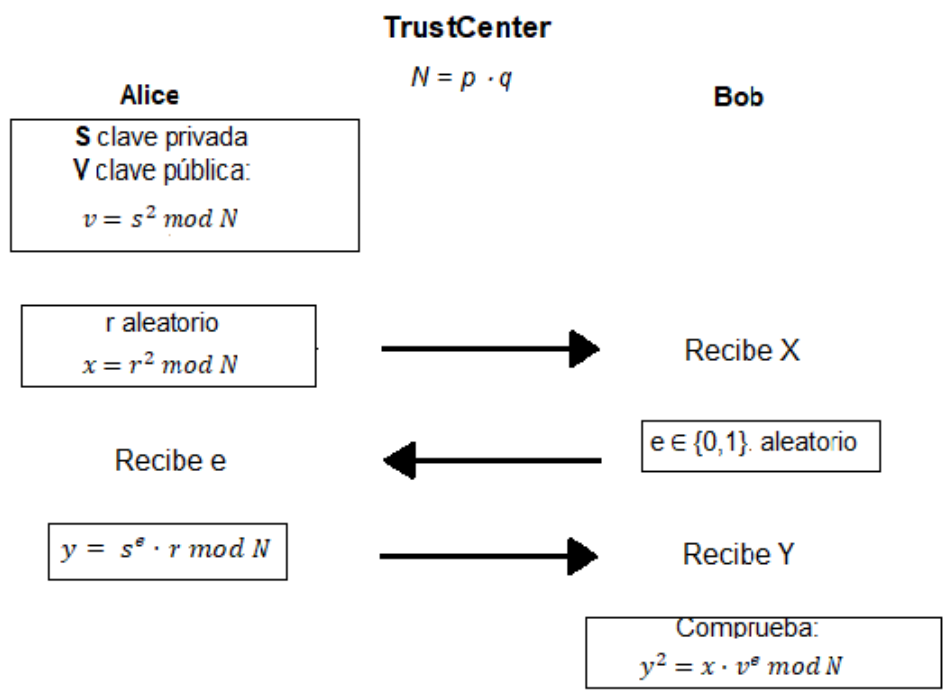


Figure 2. Fiege-Fiat-Shamir Round

For example, if we choose $e = 1$, then:

1. Public: the modulus N and $v = s^2 \ mod \ N$.
2. Alice selects r and sends to Bob $x = r^2 \ mod \ N$.
3. Bob chooses $e = 1$.
4. Alice sends $y = r \cdot s \ mod \ N$.
5. $y^2 = x \cdot v \ mod \ N$, then Bob accepts it, meaning Alice passes this iteration of the protocol.

Therefore, Alice must know s in this case.

The protocol is based on the difficulty of calculating modular square roots without knowing s. It is guaranteed by the properties

derived from the Euler's Theorem which ensure the difficulty of inverting modular operations and the appropriate dispersion of results. By repeating this process several times, Bob can be sure that Alice knows the secret s, although she never reveals it directly. Each iteration reinforces Bob's confidence that Alice truly knows the secret.

Guillou-Quisquater's Protocol

The Fiege-Fiat-Shamir's was the first practical identity-based protocol. The Guillou-Quisquater's protocol was designed to reduce the number of iterations and accreditations per iteration used in the first one by increasing computational work. For some implementations, such as smart cards, a large number of iterations is less than ideal. Exchanges with the outside world are time-consuming and the storage required for each accreditation can overload the card's limited resources.

Louis Guillou and Jean-Jacques Quisquater developed a zero-knowledge identification algorithm more suitable for applications like these. The exchanges between Alice and Bob, and the parallel accreditations in each exchange, are kept to an absolute minimum: there is only one exchange of one accreditation per proof. For the same level of security, the calculation required by Guillou-Quisquater's is greater than Fiege-Fiat-Shamir's by a factor of three. And like Fiege-Fiat-Shamir's, this identification algorithm can be converted into a digital signature algorithm (Schneier, B., 1996).

The Guillou-Quisquater's identification protocol is a ZKP scheme designed to optimize authentication in security devices (such as microprocessors or smart cards), minimizing both transmission and memory requirements, in exchange for an increase in computational work.

Once again, let n be the product of two prime numbers p and q. Let v be coprime with $\varphi(n) = (p - 1)(q - 1)$.

Let J be Alice's public key, then her private key B is defined such that $J \cdot B^v \equiv 1 \ mod \ n$.

Alice sends her credentials, J, to Bob. Now, she wants to prove to Bob that these credentials are hers. To do so, she must convince Bob that she knows B. In Figure 3, the protocol is presented:

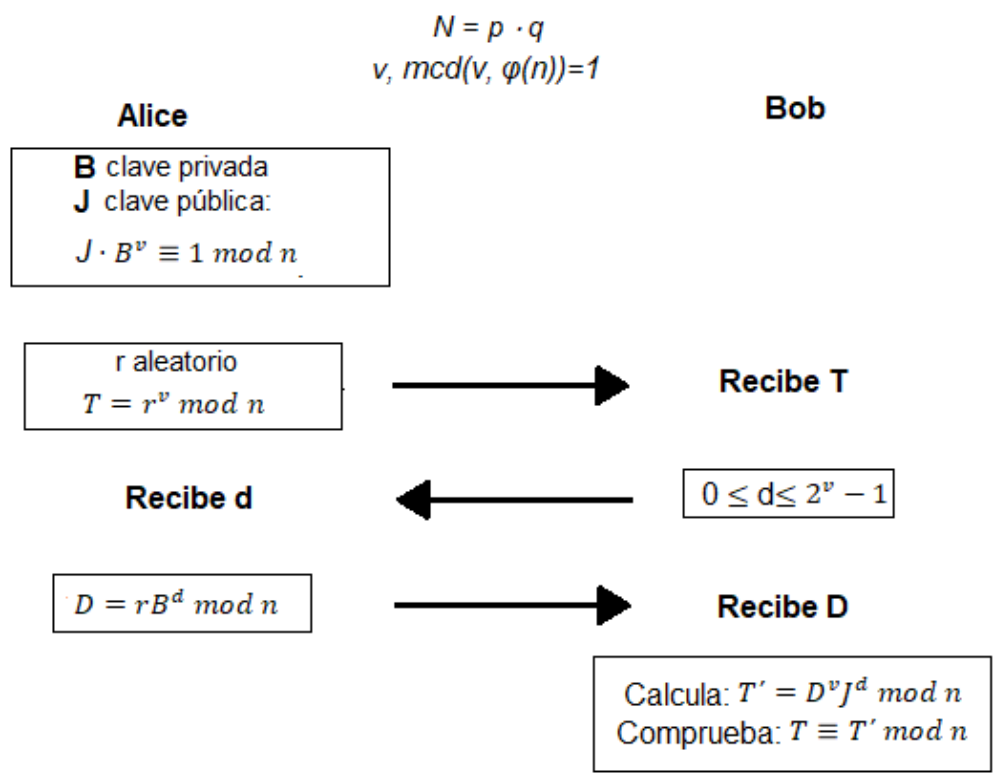


Figure 3. Guillou-Quisquater Round

1. Alice chooses a random integer *r*, such that r is between 1 and *n*-1. She calculates $T = r^v \ mod \ n$ and sends it to Bob.

2. Bob chooses a random integer, *e*, such that *e* is between zero and *v-1*. He sends *e* to Alice.

3. Alice calculates $D = rB^d \ mod \ n$ and sends it to Bob.

4. Bob calculates $T' = D^v J^d \ mod \ n$. If $T \equiv T' \ mod \ n$, then the authentication is successful.

The mathematics are not so complex:

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

Since *B* was created to satisfy $J \cdot B^v \equiv 1 \ mod \ n$ (Schneier, 1996) and (Guillou & Quisquater, 1988).

The security of this protocol lies in the parameter v, as it defines the range of values from which Bob can choose his challenge d and, consequently, the probability that an impostor could guess it. If Eve were able to guess the value of d, she could impersonate Alice even without knowing her private key s. She could choose a random value y in advance, which would be the one she sends in the third step, and define $x=y^v J^d$. In the first step, consequently, in this way, Bob's verification in the final step would be trivially true. Remember that, although Eve does not know Alice's private key, she has access to her public key (Vilchez Moya *et al.*, 2023).

Security in these protocols

One of the most common attacks on authentication methods is the replay attack. This attack involves reusing a previous successful

communication between two parties to impersonate one of them by resending exchanged messages. Although many authentication protocols, such as Kerberos, are vulnerable to this attack, ZKP protocols are resistant because each execution is unique thanks to the randomization used. In the case of interactive ZKPs, they are also deniable to third parties who did not participate in the proof, a property not maintained by non-interactive ZKPs.

Thanks to their zero-knowledge nature, security in authentication protocols based on ZKP authentication-based protocols does not deteriorate over time and is resistant to eavesdropping and chosen-text attacks. The latter attack, common in challenge-response authentication schemes, seeks to extract information about the prover's secret by selecting specific challenges. However, ZKPs prevent a dishonest verifier from obtaining information because the prover's responses always contain randomness. Furthermore, ZKPs prevent unwanted effects, such as the unintentional digital signing of a file, which can occur when the same RSA key is reused for both authentication and digital signature.

The man-in-the-middle attack, where a hacker intercepts and modifies messages between two parties, is ineffective against ZKP protocols since the secret is never transmitted over the communication channel. Additionally, ZKP protects against identity spoofing (a risk in weak password-based authentication schemes), as it never reveals the secret or related information to the verifier. Regarding phishing, ZKP offers a solid defense and prevents users from revealing their credentials through deceptive techniques such as fraudulent emails or cloned websites. Two proposed ZKP protocols provide authentication without the need to share passwords, surpassing other current schemes.

In summary, ZKP authentication protocols stand out as one of the most secure options compared to traditional schemes such as weak password authentication and strong challenge-response authentication, thanks to their resistance to numerous known attacks. Weaknesses in ZKP protocols usually arise from deficiencies in their

implementation, given that their theoretical foundations are solid and robust (Bukovits, 2023).

Conclusions

This research paper highlighted the relevance and impact of Zero Knowledge Proof (ZKP) in the field of cryptographic authentication, especially in contexts such as cloud security. The exhaustive review of the main ZKP protocols and methods has revealed that these techniques play a crucial role in improving privacy and security in distributed systems, offering robust guarantees without compromising the confidentiality of information.

The implementation of ZKP represents a significant advance in secure authentication and allows verifying the validity of a transaction or identity without needing to reveal sensitive information. The flexibility and adaptability of these methods are valuable in an environment where data protection and privacy are of utmost importance. The various applications of ZKP prove the potential of these techniques to strengthen security in diverse contexts.

Finally, privacy protection and the development of ethical frameworks for the use of authentication technologies based on ZKP are fundamental aspects that must be considered. As these techniques are integrated into broader applications, it is essential to explore approaches to guarantee data privacy and establish ethical guidelines for their implementation in sensitive environments. These areas are important for future research and development in the field of cryptographic authentication.

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/

RESEARCH

Integration of security practices for DevSecOps pipelines in containerized environments

Mary Nelsa Bonne Cuza, Eng.[*] and Meliza León Bencomo, Tech.[2]

Received: 11/2025 | Accepted: 11/2025 | Published: 12/2025

Abstract

Security in software development is a paramount concern in today's digital age, especially in continuous integration and delivery environments. As cyber threats evolve, it is necessary for development teams to integrate security from the early stages of the software lifecycle. This 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 documentation 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, the effective implementation of automated tools improves security significantly. Conclusions highlight the need to foster a collaborative culture and a proactive approach to security. The urgency of addressing security concerns in software development is highlighted and a clear process towards successful integration of DevSecOps practices is proposed.

Keywords: containerization, continuous delivery, continuous integration, security

[*] Faculty of Cybersecurity. University of Computer Sciences. Carretera a San Antonio Km 2 ½ Torrens. Boyeros. La Habana. Cuba. mary@uci.cu

[2] Directorate of Computer Security. University of Computer Sciences. Carretera a San Antonio Km 2 ½ Torrens. Boyeros. La Habana. Cuba. mary@uci.cu

Introduction

The evolution of software development practices has led to the creation of methodologies aimed at improving efficiency and collaboration among the involved teams. One of the most significant innovations in the field is the concept of DevSecOps, which has transformed the way in which developers and operation teams interact and work together. Its purpose is to optimize the software delivery process. This methodology promotes collaboration between operators and developers, with the objective of aligning their agility towards a common goal (Muñoz, 2023).

Information security and regulatory compliance are essential in software development. Recent statistics reveal a prominent increase in the costs associated with data breaches, as well as in the frequency and sophistication of attacks targeting applications and systems. Vulnerability management has become an urgent challenge, with remediation times exceeding 200 days, according to data from the National Institute of Standards and Technology (NIST) (Madnick, 2023).

The scenario is marked by an increasing focus on automation and ongoing enhancement of software operations. Between 2015 and 2025, a significant adoption of DevSecOps has been observed; whose exponential growth reflects the progressive maturity of integrated security practices. In this context, DevSecOps is consolidated as the key approach to address security challenges. The methodology is established as an imperative necessity to mitigate risks and ensure the integrity of applications (Chandramouli *et al.*, 2024; GitLab, 2025). Technologies such as Docker and Kubernetes are used to improve deployment speed, patching, scalability, and efficiency in software development. In this research, the authors focused on the Kubernetes containerized technology.

Kubernetes enables the effective work's organization and management in DevSecOps by providing a scalable and automated infrastructure that simplifies application deployment. Many organizations have adopted this technology to implement specific security policies and reduce the attack surface, improving resilience against potential

threats. The integration of DevSecOps in Kubernetes is materialized through a continuous security pipeline that encompasses multiple protection layers (Kubernetes, 2024).

However, the environment shows a sequence of critical challenges related to pipeline security. One of the most significant problems is revealed when a recent deployment is compromised by an undetected vulnerability in one of the container images. The absence of specific standards and established best practices prevents the development team from having a reference framework to guide the implementation of security measures throughout the SDLC (System Development Life Cycle).

This lack of regulation delays the detection of problems and compromises the application's integrity, which negatively affects customer trust and corporate reputation. The situation generates uncertainty regarding regulatory compliance that could lead to significant financial penalties and the loss of operating licenses for the business. The lack of visibility into the pipeline status and performance also represents a critical challenge. Without appropriate monitoring tools, it is nearly impossible to detect anomalies or suspicious performance in real-time. This situation leads to an increase in incident response times, severely impacting operational continuity. In the absence of standardized practices for managing secure configurations, the development team identifies that many applications are deployed with insecure default configurations, generating an environment prone to cyberattacks and vulnerabilities susceptible to being exploited by malicious actors.

Based on the foregoing, the authors decided to address this research on the most suitable security practices that can be integrated into the DevSecOps pipeline in containerized environments to guarantee comprehensive software protection from its conception to its deployment in production.

Materials and Methods

To develop the proposed solution, the authors considered various scientific methods that allowed for a greater organization in their work.

The analysis-synthesis method was used to break down the research problem into its main components and subsequently integrate these elements into the proposed procedure. This process enabled the identification of interrelationships among the tools, standards, and security best practices used. The observation method contributed to defining the object of study and the tools employed; it enabled the collection of relevant information on the research topic. The induction-deduction method was used to gain a clear vision of the procedure to be developed.

The necessary basics regarding the elements involved in DevSecOps pipelines in Kubernetes were taken into account, such as continuous integration and delivery (CI/CD) in the implementation of the DevSecOps pipeline, which ensures that each stage of the software lifecycle is supported by robust measures (Muñoz, 2023; Xygeni, 2024). According to Kev Zettler (in his publication for Atlassian, 2024), the implementation of DevSecOps has five cyclical stages for its development: Planning, Compilation, Testing, Deployment and Monitoring (Atlassian, 2024). These stages become an essential framework to ensure that applications are secure and aligned with the organization's strategic objectives. Meanwhile, Luca Antoniotti, a software developer, defines, in his master's thesis for the Polytechnic University of Turin, four fundamental stages for the DevSecOps pipeline with subtle differences from the general methodology: Building, Testing, Releasing, and Deployment (Antoniotti, 2021).

Standards and regulations for DevSecOps were considered. Among the main ones are:

- ISO/IEC 27001:2022: provides an adaptable framework for aligning security with business objectives (Zhu & Zou, 2021).
- NIST SP 800-53: provides specific guidelines for protecting systems and data in dynamic environments, such as CI/CD pipelines (Zayni & Nassereddine, 2020).

An analysis of solutions for integrating security standards and best practices for the DevSecOps pipeline in Kubernetes was carried out based on the design of a search protocol that enabled the observa-

tion of papers and studies related to the research topic and laid the foundations for the proposed solution development.

This solution integrates tools and technologies that establish security standards and best practices for the DevSecOps pipeline in containerized environments. Tekton was used, an open-source automation framework (specifically within Kubernetes environments), particularly to create CI/CD systems, which allows developers to build, test, and deploy on cloud providers and local systems. The choice of Tekton is justified by its ability to provide an extensible and modular platform that enables workflow automation (Tekton, 2023).

Results and Discussion

In the field of process management and security controls, structuring defined procedures is fundamental to ensuring consistency and compliance with established objectives. The procedure proposed in this research consists of three main phases, which group a total of six steps that allow for systematically defining, implementing, and validating security controls, ensuring their alignment with recognized standards and the correct assignment of responsibilities. Figure 1 shows the process organized into three main phases:

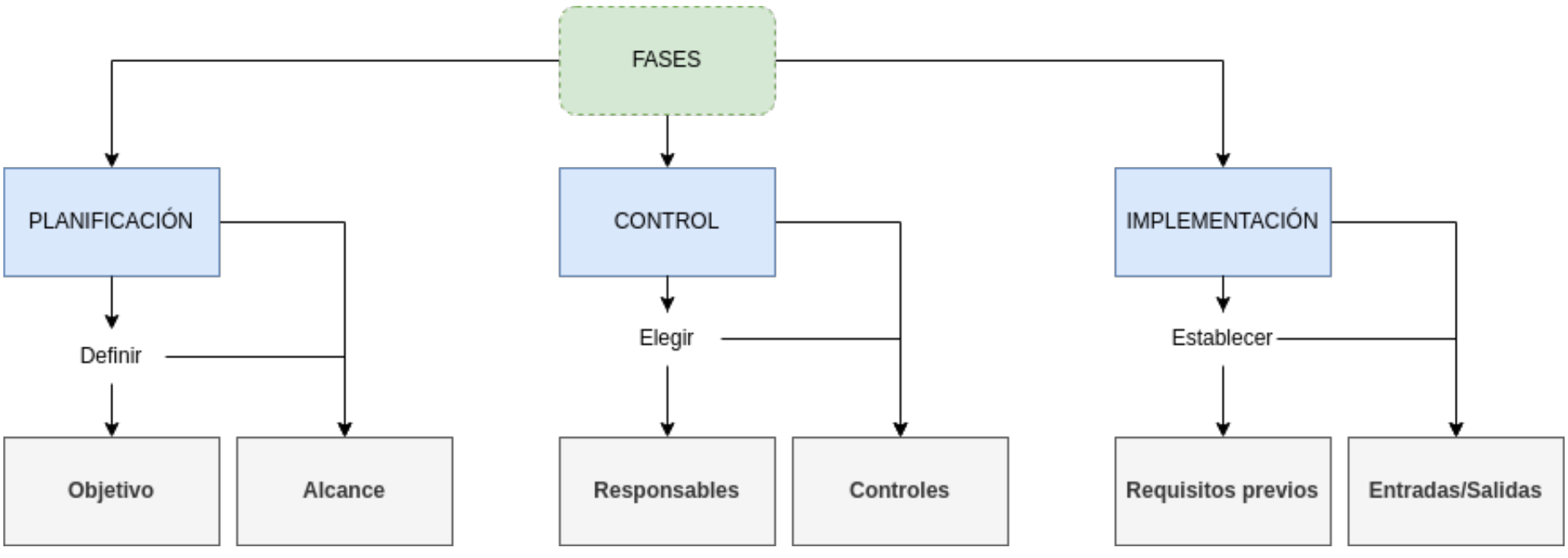


Figure 1. Structure of the procedure to be developed

1. Planning Phase

- It begins with the objective definition, which establishes the main purpose.
- The scope is derived from the objective and defines the procedure boundaries.

2. **Control Phase**

- The selection of security controls provides the necessary measures.
- The assignment of responsible parties ensures proper execution.

3. **Implementation Phase**

- Specific requirements for each control are established.
- Finally, the inputs and outputs adapted to the process are defined.

Through this approach, the aim is not only to establish a defined action framework but also to enable adaptability to different organizational contexts, integrating prerequisites, expected actions, and feedback mechanisms. The methodological application of these steps contributes to continuous improvement, minimizes risks, and optimizes the effectiveness of the implemented controls.

Description of the Phases for the Procedure Implementation

1. The **planning phase** is fundamental to any procedure. Its main objective is to clearly and precisely establish what is desired. This process provides a specific direction and defines a tangible purpose that will guide all subsequent actions (see Table 1).

Table 1. Procedure Planning Phase

Procedure Objective	Procedure Scope
• To integrate automated security controls into the CI/CD pipeline using recognized standards (NIST SP 800-53 and ISO/IEC 27001:2022), which combines technical implementation. • The unified approach seeks to ensure that security is not a later phase, but an automated process from build to deployment, reducing risks without compromising agility	• Multidimensional Scope: The proposed solution impacts three key areas: the business domain, spanning all organizational levels from development to operations; the technical domain, covering the Kubernetes infrastructure and its components; and the security domain, implementing controls to ensure compliance with international standards, such as ISO/IEC 27001:2022 and NIST SP 800-53.

2. The main purpose of the **Control Phase** is to define a set of security and verification measures that ensure rigorous compliance with established standards. It involves the selection of mechanisms to monitor work quality and the designation of specific individuals responsible for assigned tasks. This ensures that each team member has a defined role to maintain the process integrity.

Standards and Controls. Selection Criteria

The choice of *ISO/IEC 27001:2022* together with *NIST SP 800-53* as security frameworks for a DevSecOps pipeline is justified due to the comprehensive coverage and complementary approach they respectively possess. The regulatory basis for the security integration procedure in DevSecOps pipelines includes the following specific elements, each one with a generalized function, according to the regulations (Wilbur L. Ross, Jr., Secretary, 2020; *ISO/IEC 27001, 2022*).

ISO/IEC 27001/2022

- A.8.1 It focuses on managing and securing end-user devices.
- A.8.2: it establishes requirements to manage and control privileged access within the organization.
- A.12.1.3: it establishes the necessary components to maintain adequate capacity to ensure services are secure and available.
- A.8.23: it implements measures to control and monitor access to web content.
- A.8.28: it focuses on ensuring that software development is carried out securely.

NIST SP 800-53

- AC-2.7: it establishes a unique and consistent authentication framework, allowing automatic integration into the pipeline through automation of authentication testing and continuous identity validation. It establishes access control to resources.
- AC-3: it provides an additional security layer through access management, which is essential in DevSecOps environments where changes are frequent and need to be consistent.

- AC-6: it provides a minimal privilege management approach, enabling the automation of roles and permissions.
- AC-4: it establishes secure network policies that can be versioned and automatically applied, enabling network segmentation and traffic control at each stage of the pipeline.
- AC-10: it provides a comprehensive security event management approach that can be automated and monitored in real-time.

These controls are particularly valuable because they can be automated and versioned, allowing for their seamless integration into the CI/CD pipeline, unlike other controls that require manual intervention or are not directly applicable to an automated environment. Their joint implementation creates a comprehensive security layer that covers the main aspects of a DevSecOps environment: authentication, access control, network security, and event monitoring.

Selection Strategy for Security Controls

For this research development, a selection protocol was designed to enable the analysis of standards and controls related to the integration of security measures in the DevSecOps pipeline within the k8s cluster using Tekton. The combination of these criteria allowed for the selection of a set of controls that not only meet security requirements but are also implementable and maintainable in a modern DevSecOps environment, where automation and delivery speed are fundamental.

Purpose of the Analysis: To identify standards and controls that describe applicable methods for integrating best practices for the DevSecOps pipeline in Kubernetes. Additionally, to identify secure implementation techniques that can be integrated into the DevSecOps pipeline.

Information Source: Government regulations or those verified at the national level. *ISO/IEC 27001* y *NIST SP 800-53*.

Search Strategy: Controls that cover the implementation of security practices that complement the robustness of Kubernetes.

Inclusion and Exclusion Criteria: Inclusion: Automation capability through the possibility of implementation via scripts.

Alignment with security measures established in the Cuban regulatory framework. Integration with used automation tools. Ability to evolve with the organization based on its adaptability to different business environments.

Review Procedures: Initial selection, detailed evaluation, and synthesis of results.

Parties Responsible for Procedure Execution: The DevSecOps approach transcends development, operations, and security teams. It involves multidisciplinary groups including Quality Assurance (QA) teams, solutions architects, and standard compliance specialists. Collaboration among these groups is fundamental to ensuring the integration of security into every phase of the pipeline development lifecycle.

The DevSecOps team is responsible for the process. Its main function is to ensure the integrity and protection of development and deployment environments. Its responsibilities include the definition and maintenance of security policies, based on NIST SP 800-53 standards. This team is in charge of implementing automated controls on the Tekton platform, in compliance with the selected standards.

End Customers: The process has various customers divided into two groups: internal, who actively participate in the DevSecOps process; and external, who are the final beneficiaries with a relationship independent of the process. Pipeline implementation in DevSecOps is fundamental because it automatically integrates security into every stage of software development, helping to identify and correct failures from the beginning of the process. By automating security testing and vulnerability assessment, organizations can deliver more secure, higher-quality software faster.

Prerequisites for Implementing of Each Control: Securely organizing and isolating resources allows for the creation of a solid foundation for developing and deploying the selected controls. The following are the requirements that ensure a successful implementation for creating and managing pipelines:

Requisite 1: Install and Configure Tekton: It must begin by executing the instructions corresponding to its initial documentation. The

objective is to ensure that the code and dependencies comply with security standards before generating artifacts.

Requisite 2: Configure the working environment

Expected Outputs: An operational DevSecOps pipeline that integrates automated security controls, establishing a continuous feedback loop for the iterative improvement of security measures throughout all stages, from development to production.

Implementation in the Pipeline Building Stage

The building stage is essential for verifying code integrity and identifying security vulnerabilities. Its development is carried out through *NIST 800-53* controls' AC-3 and AC-2.7.

NIST SP 800-53 Control AC-2(7) - Role-Based Access Control (RBAC)

Objective: Restrict access to the building pipeline (e.g., only the DevSecOps team can modify *Tasks*).

Implementation

- Create 3 YAML Roles to define essential roles (a full admin role, a developer role, and a read-only role).

NIST SP 800-53 Control AC-3 - (Network Policies)

Objective: Ensure that only authorized resources can interact with the pipeline, applying the principle of minimal privilege and security by default. Create NetworkPolicies to isolate pods.

Implementation

- Create NetworkPolicies to isolate the pipeline.

Implementation in the Pipeline Testing Stage

The implementation of I*SO/IEC 27001:2022* as a security control in the DevSecOps pipeline was carried out through an integrated design that began with the definition of two fundamental controls: A.8.1 for identifying assets through secret scanning, in accordance with Cuban legislation recommending that entities implement security measures for critical information according to Article 17 of *Decree-Law 370/2018*; and A.8.2 for controlling elevated access to systems and data in dependencies, following the provisions of *Decree-Law 360/2019*, which requires risk management in critical infrastructures according to its Article 8 (Saber, 2018).

ISO/IEC 27001-2022 Control A.8.1 and A.8.2 (Access and Privilege Management)

Objectives: Strengthen security during application development and deployment through the controlled assignment of administrative rights and scanning for private or restricted-access elements.

Implementation:

- Create *Task* for the scanner.

The task is designed to scan secrets in a source code directory using the Aqua Security Trivy tool, which helps to identify vulnerabilities and security issues in different components.

Implementation in the Pipeline Delivery Stage

To establish a secure environment in Tekton Pipelines during the delivery stage, various aspects of different *NIST SP 800-53* controls were implemented, in compliance with the Cuban legal framework that establishes security controls for state entities, including access management and data protection in *Resolution 105/2020*.

NIST SP 800-53 Controls AC-3, AC-6, AC-4, AC-10

Implementation

- Create a namespace and a service account with minimum permissions.

A ServiceAccount is created within the previously created namespace. The service account will be used by the pods to access necessary resources within the cluster. Establishing this service account facilitates the execution of automated tasks and the implementation of continuous integration processes.

- Apply to the cluster with *kubectl* apply -f <service-account.yaml>
- Define a role that only allows executing pipelines without administrative permissions.

Implementation in the Pipeline Deployment Stage

In Kubernetes environments for deployment, multiple technologies and tools (such as Docker, Helm, Istio, among others) are used, which may vary among organizations. At this stage, security best practices are proposed, according to the necessary characteristics of each project or software deployed based on the

organization. Recommending general standards allows each team to choose the most appropriate tools for their context without being limited to a specific solution that may not be inclusive for their needs. The deployment stage in a DevSecOps pipeline is critical for ensuring operational security, where specific controls from the *ISO/IEC 27001:2022* standard will be applied. The selected controls cover A.12.1.3 (capacity management), A.8.23 (network services security), and A.8.28 (protection for public systems).

ISO/IEC 27001:2022 Control A.12.3 (Capacity Management)

The control establishes that resource usage must be monitored, adjusted, and projected to anticipate future needs, thus ensuring the required system performance. It also emphasizes the importance of scanning and evaluating system components to identify known vulnerabilities. To implement it, it is recommended to add steps in the Deployment Task that fulfill the proposed functions.

Objective: Limit CPU per pod to avoid excessive consumption.

ISO/IEC 27001:2022 Control A.8.23 (Network Security)

The control focuses on ensuring that communications within systems are protected against unauthorized access, attacks, and exposure of sensitive data.

Objective: Restrict traffic among pods (e.g., only frontend and backend).

ISO/IEC 27001:2022 Control A.8.28 (Protection for Public Systems)

The control refers to configuration and change management in information systems. It establishes that system configurations must be documented, controlled, and regularly reviewed to ensure they do not introduce vulnerabilities. Furthermore, configuration changes must be evaluated, tested, and approved before implementation to minimize risks.

Objective: Ensure communications with HTTPS (Hyper Text Transfer Protocol Secure) and avoid data in plaintext. Force HTTPS in Ingress by redirecting HTTP to HTTPS and using HSTS.

Validation Strategy for the Proposed Procedure

Procedure validation is essential in the field of computer security, as it is a rigorous process that ensures established methods effectively meet the proposed objectives. Through this procedure, the capability of each element is verified, potential failures are identified and corrected, proving its suitability to protect computer assets. For this research, two approaches were used: simulation and metrics.

Figure 2 details the procedure for developing the VirtualBox simulation. This guide describes each stage of the process in order to ensure the correct techniques' application and achieve successful validation in a real environment.

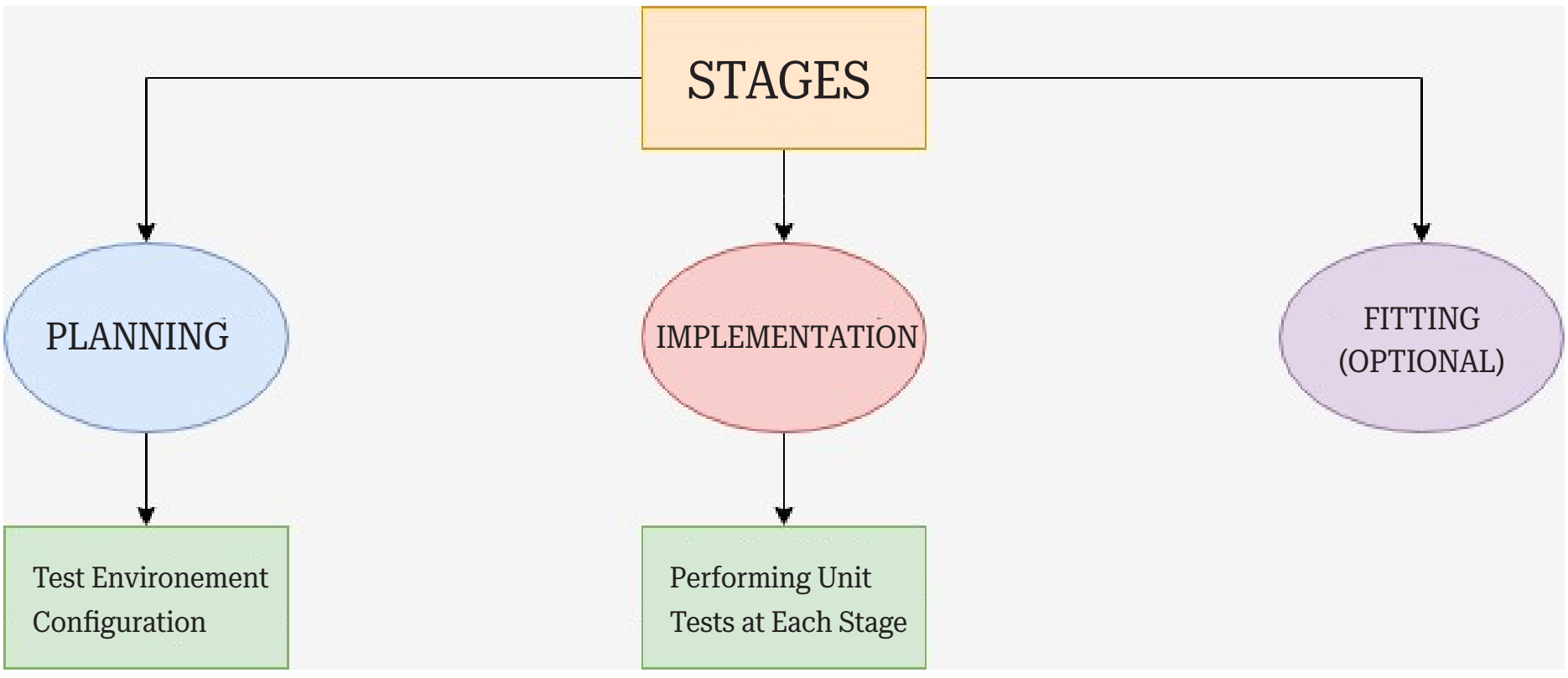


Figure 2. Stages comprising the process used in the simulation validation strategy

The two main stages, join to an optional phase, simplify the establishment of clear objectives, the identification of necessary resources, and the success criteria's definition. The fitting stage includes the correction of identified findings, configurations' adjustment based on results, identification of areas for improvement, documentation of implemented changes, and prioritization of corrective actions based on business needs.

The stages materialize the procedure into concrete actions. This enables the controlled reproduction of security scenarios and the practical verification of procedures. Furthermore, it contributes to the analysis of the results obtained and generates relevant conclusions that serve as a basis for recommendations aimed at continuous

improvement. The logical sequence of stages ensures that the simulation is a rigorous process, in which each phase builds upon the previous one to produce reliable and applicable results.

The use of **metrics** to determine regulatory compliance is essential to ensure that the organization complies with applicable security standards and regulations, facilitating the identification of gaps and the continuous improvement of security policies. Maverix is a key solution for implementing DevSecOps due to its ability to automate and centralize critical security metrics, integrating different tools into CI/CD pipelines. Its quantifiable model allows for prioritizing vulnerabilities, reducing risks, and demonstrating regulatory compliance, while fostering an agile security culture (Maverix, 2021).

Maverix calculates Compliance with Policy (CP) using the formula $CP = \frac{PC}{PT} \times 100$, where PC is the number of implemented security checkpoints and PT is the total number of controls required according to recognized standards. This metric allows for evaluating security gaps (e.g., an 80% CP indicates 20% pending controls) and prioritizing their implement ation to comply with regulations or improve the security posture. The acceptable compliance threshold is established when CP is greater than or equal to 75%.

The 75% compliance threshold was selected as acceptable, in line with recommendations from various international frameworks:

- According to section 6.1.2 of the *ISO/IEC 27001* standard, security controls can be adjusted based on the level of risk.
- The NIST CSF framework classifies maturity levels into four categories (Tier 1 to Tier 4), where Tier 3 ("Consistent") indicates that most relevant controls are in practice. 75% compliance may correspond to a Tier 3 maturity level, meaning the organization has a stable security posture but has not yet reached excellence, which would be Tier 4.
- The CIS Controls (a set of best practices designed to help organizations improve their cybersecurity) prioritize basic and foundational controls. For low-criticality systems, implementing at least 75% of the basic controls is recommended.

From an economic perspective, achieving 100% compliance would imply a substantial increase in costs, without offering proportional benefits in risk reduction. Therefore, it is concluded that a compliance level between 75% and 80% is common in similar environments (test or low criticality), supporting the choice of this threshold as a reasonable and practical option.

Conclusions

As a result of the conducted research, it is concluded that:

Understanding the vital concepts of DevSecOps is crucial for implementing effective security practices in software development. Detailed analysis of standards and best practices allows for establishing a solid theoretical-methodological framework that supports the successful implementation of security procedures, guaranteeing the integrity and reliability of the development process.

Implementing security standards and best practices in DevSecOps pipelines is essential to protecting business infrastructure and ensuring software integrity. The structured procedure provided allows for identifying and mitigating vulnerabilities to strengthen the organization's security through a systematic and controlled approach.

The validation of the proposal consolidates its credibility through a reflective approach that allows for adjusting results based on the data obtained, contributing to the scientific knowledge’s advancement. The conducted simulations confirmed that the workflow is robust and that the procedure achieves its intended objectives. This ensures the procedure's practical applicability and ability to generate improvements.

Integrating security standards and best practices into the DevSecOps pipeline is essential to proactively protect software from the initial stages of development. It enables identifying and addressing vulnerabilities before they become costly problems, ensuring regulatory compliance and improving software quality.

References

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

RESEARCH

Deployment of Honeypots for Proactive Cyber Threats Detection

Heidy Rodríguez Malvarez, Eng.[*], Elizabeth Molina Mena, Eng.[2], Mónica Peña Casanova, Ph.D.[3]

Received: 11/2025 | Accepted: 11/2025 | Published: 11/2025

Abstract

In an increasingly threatened digital environment by malicious players, the need to implement proactive security mechanisms has become crucial. This paper 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. First, the state of honeypots and their applications in cybersecurity are addressed, emphasizing their relevance as a deception technique to observe and analyze attacker behavior. Second, a flexible and scalable architecture implemented in a controlled laboratory environment, where various attacks are simulated to evaluate system effectiveness, is defined. The obtained 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.

[*] Department of Technology Infrastructure, School of Cybersecurity, University of Computer Sciences. heidyrm@uci.com

[2] Department of Information Security, University of Computer Sciences. angelagv@uci.cu

[3] Technology Infrastructure Office, Dean of the School of Cybersecurity, University of Computer Sciences. monica@uci.cu

Keywords: cyber threats, cybersecurity, proactive detection, honeypot, T-Pot

Introduction

In the contemporary cybersecurity environment, organizations face a growing sophistication and diversification of cyber threats (Fadziso *et al.*, 2023). Malicious actors, ranging from isolated individuals to organized groups, use increasingly advanced tactics aimed at compromising computer systems, stealing confidential information, or disrupting essential services (Cherqi et al., 2023). This situation has driven the need to adopt proactive mechanisms that make it possible to anticipate attacks and respond effectively before they cause irreversible damage.

Traditionally, the defense strategies of security operations centers (SOCs) have been based on the use of tools such as intrusion detection and prevention systems (IDS/IPS) and antivirus programs. However, these methods have limitations when they face the evolution of evasion techniques employed by attackers, who manage to circumvent solutions that rely on previously known signatures or patterns (Skopik *et al.*, 2016). Hence, it is imperative to implement solutions that go beyond reactive detection and enable the observation and analysis of actual attacker behavior in a controlled environment.

In this context, honeypots emerge as an effective technology for identifying and studying malicious activities. These systems, designed to appear vulnerable and attractive to attackers, act as decoys and make it possible to record their tactics, techniques, and procedures without jeopardizing the organization's actual assets (Lorusso Montiel & Uc Ríos, 2022). The obtained information through honeypots proves invaluable for strengthening technological infrastructure defenses and anticipating future threats.

This research aims to evaluate the deployment of honeypots as a proactive measure for the early detection of cyber threats in organizational environments, with special emphasis on the use of the

T-Pot tool, recognized for its modular architecture, integration capacity, and support for multiple emulation services.

To achieve this objective, a solution based on a controlled laboratory environment was implemented to simulate different types of attacks and to analyze the effectiveness of the tool in capturing, monitoring, and correlating data related to security incidents. This approach contributes, not only to the experimental validation of the honeypot but also to the strengthening of organizations' cyber defense strategies.

Materials and Methods

The analysis of this research was based on a mixed methodological approach. Theoretical and empirical methods were combined to design, implement, and validate a honeypot architecture aimed at the proactive cyber threats detection in organizational environments.

General Research Design

The research was structured into three fundamental phases: design, implementation, and validation of the T-Pot honeypot.

In the first phase, a literature review was accomplished on proactive threat detection mechanisms and the different types of honeypots, which enabled the selection of the most suitable tool for the requirements of the Empresa de Tecnologías de la Información , (hereinafter, ETI, for its Spanish acronyms) of BioCubaFarma.

In the second phase, an experimental architecture was designed based on the deployment of T-Pot in a controlled laboratory environment, where network and web attacks were simulated to reproduce similar conditions to those faced by a security operations center (SOC).

In the third phase, empirical validation of the system was carried out through the use of detection metrics, event correlation, and alert analysis. This methodological structure demonstrated the effectiveness of T-Pot as a complementary mechanism to the traditional security solutions (IDS/IPS and antivirus) used by ETI 2024.

Environment and Technological Tools Used

The experimental environment was deployed in a virtual infrastructure consisting of:

- Main server (Ubuntu Server 22.04 LTS): the T-Pot platform, an open-source honeypot solution developed by Deutsche Telekom, was implemented.
- Attack virtual machines: these were used to simulate threats using tools such as Nmap and OWASP ZAP, running on a Kali Linux operating system.
- Controlled virtual network (VirtualBox): configured to isolate the experiments from the production environment and ensure laboratory security.

Figure 1 shows the environment described above:

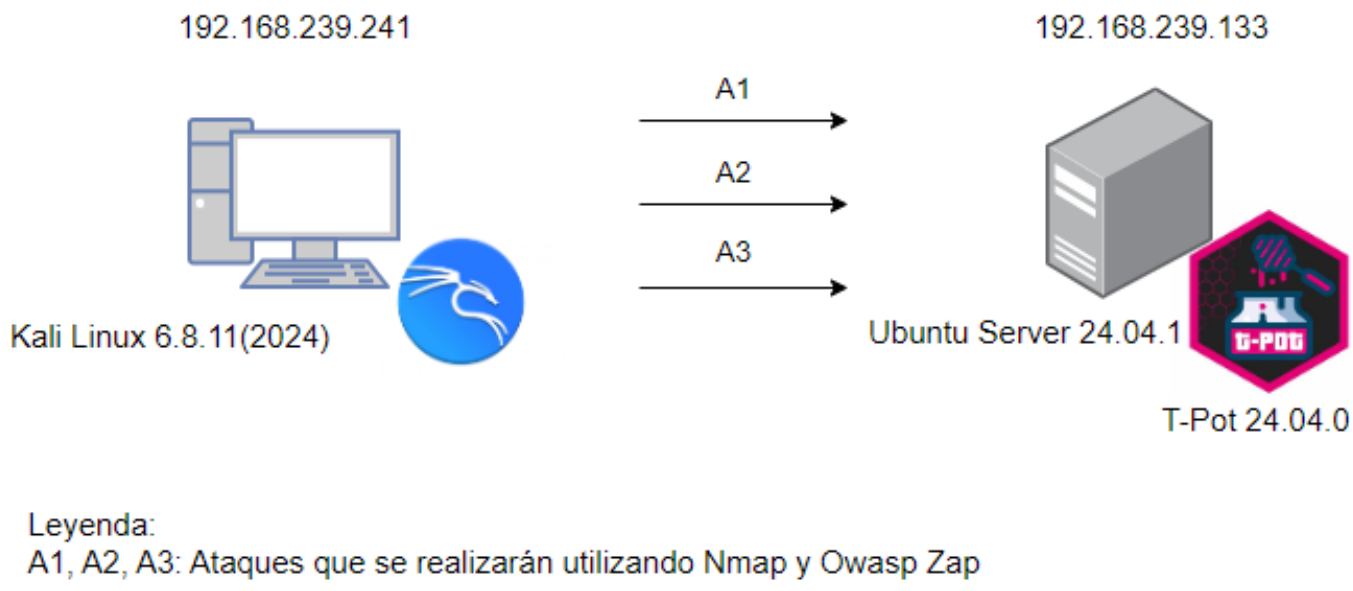


Figure 1. Controlled laboratory environment

The T-Pot honeypot integrates various service emulation modules, among which Cowrie (SSH/Telnet), Dionaea (malware catcher), Elasticpot (Elasticsearch honeypot), Heralding (authentication server), and Mailoney (fake SMTP) stand out. This modular architecture facilitates the collection and classification of events according to the attack vector and the target service (Telekom Security, 2024).

For data monitoring and analysis, complementary tools were employed, such as:

- Kibana: it enables the visualization of metrics and attack indicators.
- Elasticsearch: it functions as the central storage and analysis engine.
- Suricata: it is used for the detection and correlation of malicious traffic.

- Spiderfoot y Elasticvue: it is used for automated hosts reconnaissance and queries to the captured data indices.

Scientific Methods

The following theoretical and empirical methods were applied:

- **Historical-logical:** to analyze the evolution of cyber threats and the progression of digital defense strategies.Analytical-synthetic: to study existing honeypot technologies, identify their advantages and disadvantages, and substantiate the choice of T-Pot as the experimental solution.
- **Modeling:** to design the simulated network environment that replicated the technological infrastructure of the IT Company (ETI), in order to evaluate the tool's effectiveness under real attack conditions.
- **Measurement:** to assess the honeypot's performance based on the number and type of events detected, frequency of generated alerts, and effectiveness in data correlation.

These tools and methods provided a solid experimental framework, characterized by a controlled honeypot configuration that ensured the reproducibility of the study and the validity of the obtained results.

Results and Discussion

During the validation phase, the T-Pot honeypot demonstrated its ability to capture and analyze a wide range of malicious activities in real time. The system recorded events related to network scans, unauthorized authentication attempts, denial-of-service attacks, and vulnerability exploitation in simulated services.

Figure 2 shows a highly active environment in terms of cyber threats, with a total of 34,818 attacks directed at multiple honeypots. From this total number of attacks, it can be concluded that Honeytrap was the honeypot that detected the most attacks, with a total of 32,507 attacks, indicating the prevalence of automated scans and unsolicited connections targeting common services.

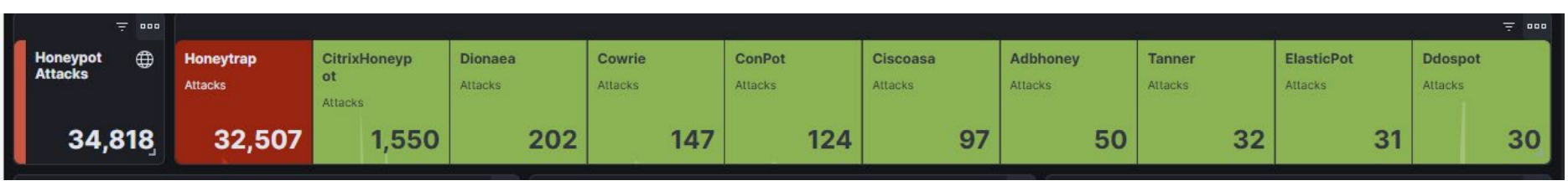


Figure 2. Attacks Detected by T-Pot

In the conducted tests, it was observed that the most attacked services were SSH (port 22), HTTP (port 80), and SMB (port 445), evidencing the prevalence of automated reconnaissance and exploitation tactics. These activities were visualized through Kibana dashboards, where histograms, density charts, and geographic maps representing the origin and volume of attacks were generated.

The Suricata engine generated multiple alerts classified according to severity and incident type, as shown in Figure 3. The predominant category is *Generic Protocol Command Decode*, represented in red, which shows a very high volume of suspicious activity related to generic protocol commands, associated with reconnaissance or mass scanning attempts. To a lesser extent, other categories were recorded, such as *Potentially Bad Traffic* (green), which includes traffic with anomalous characteristics; *Web Application Attack* (blue), representing attack attempts targeting web applications; and A *Network Trojan was Detected* (purple), related to Trojan patterns. Additionally, traffic marked as *Not Suspicious Traffic* (light pink) was identified, which, although not clearly exhibiting malicious behavior, was recorded for analysis.

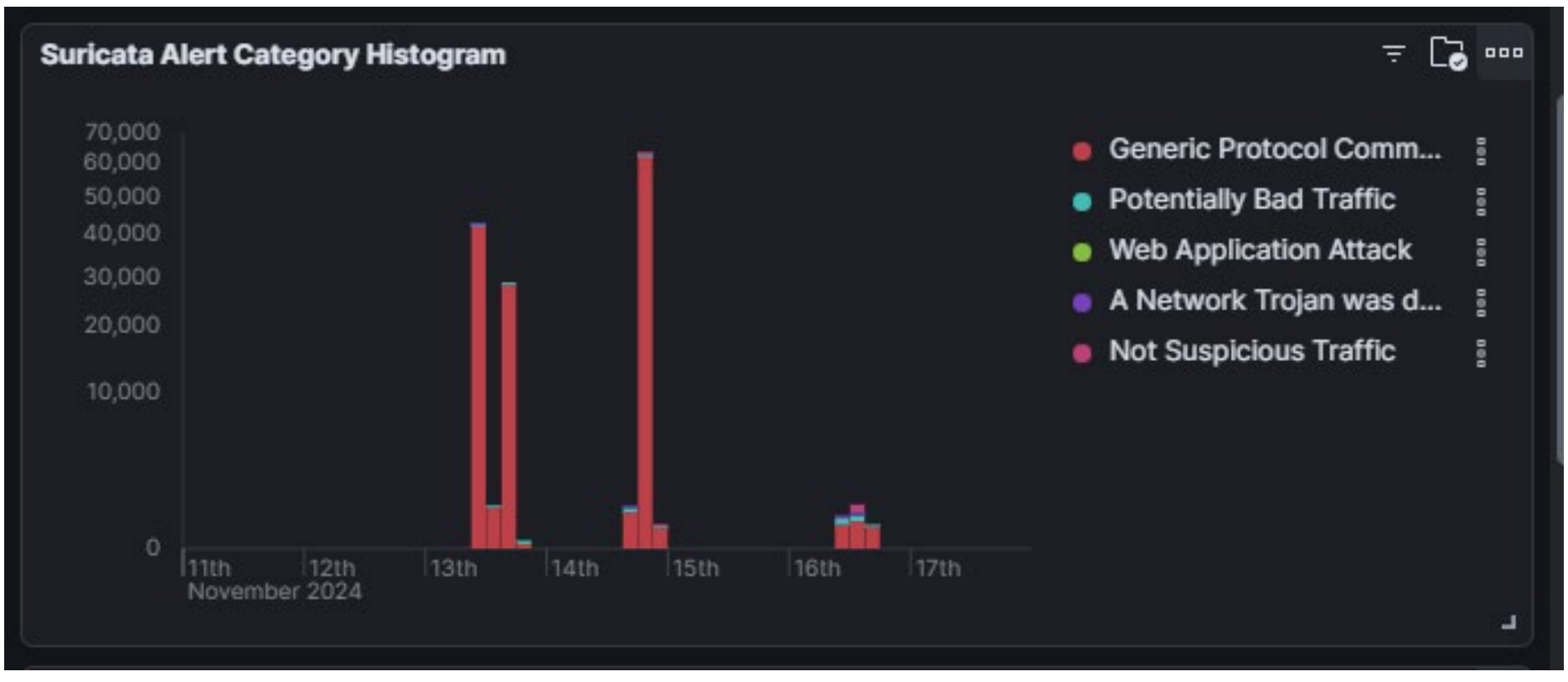


Figure 3. Alerts Generated by Suricata

One of the most relevant findings was the identification of repetitive patterns in attack tactics. Figure 4 presents tag clouds visualizing information related to authentication attempts in the traffic captured by the honeypot. In the username tag cloud, several text entries corresponding to the usernames used in the requests are

observed. The word "anonymous" stands out the most, indicating that many access attempts were made with an anonymous or unidentified username. Other entries, such as "sip" or "From: sip:nm@nm", could suggest access attempts targeting services related to the SIP (*Session Initiation Protocol*), which is very commonly used in VoIP communications. Other terms such as "Max-Forwards: 70" and "Call-ID: 50000" reinforce the hypothesis of an attack or scan directed at VoIP services. Meanwhile, the password tag cloud highlights the password attempts observed during the attacks. Some of the most common passwords were "anonymous", "(empty)", "IEUser@", and "lolo", indicating that attackers attempted to gain access using weak or easily guessable credentials. The term "(empty)" implies that attempts were made to initiate sessions without a password, while "anonymous" reinforces the use of unauthenticated access.

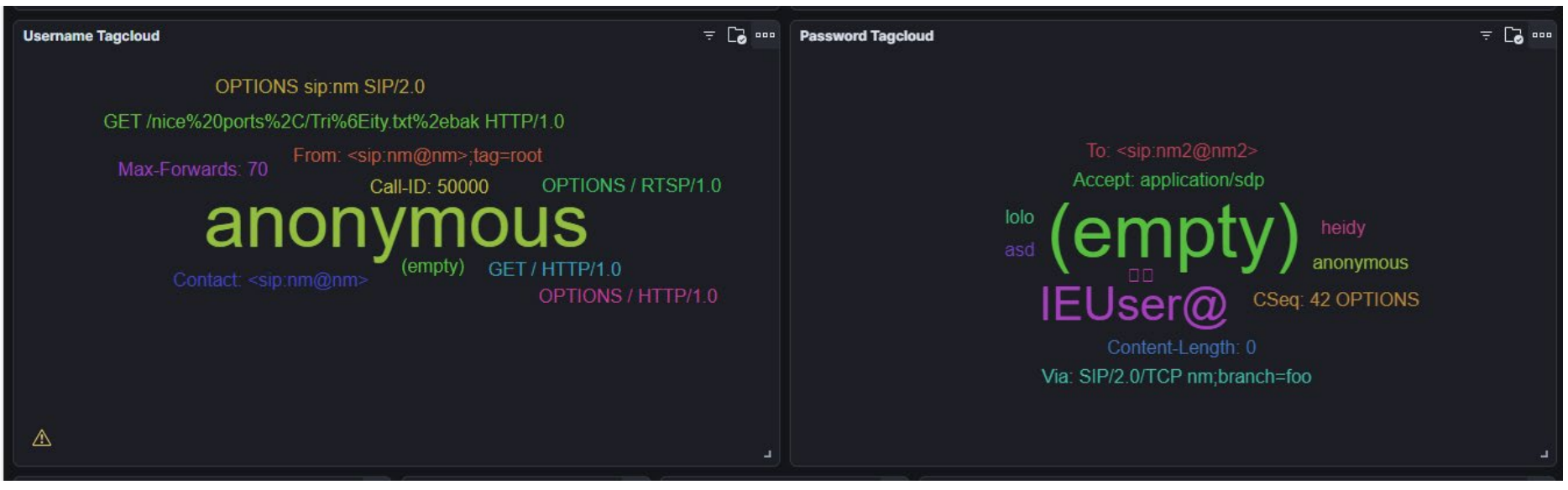


Figura 4. Nube de palabras

The obtained results demonstrate that the deployment of T-Pot is not only useful for early threat detection but also contributes to cyber threat intelligence (CTI), enabling organizations to observe the complete attack cycle, from the reconnaissance phase to execution.

Furthermore, the use of integrated tools such as *Elasticvue* facilitated precise queries on the collected data, optimizing forensic analysis and incident classification. This integration highlights the strength of T-Pot as a scalable and adaptable platform, capable of being incorporated into corporate or academic infrastructures for monitoring, training, and threat analysis purposes.

Conclusions

The deployment of the T-Pot system as an advanced honeypot in a controlled laboratory environment validated its effectiveness as a proactive mechanism for detecting cyber threats. The results confirm that this tool is capable of identifying intrusion attempts, network scans, and malicious behavior patterns, offering a comprehensive view of the threat landscape.

T-Pot's modular architecture, in conjunction with its native integration with analysis tools such as Kibana, Elasticsearch, and Suricata, provides an efficient environment for the collection, visualization, and correlation of events in real time. This capability represents a significant advantage over traditional mechanisms by enabling an early response to emerging threats.

Additionally, the research shows that the adoption of honeypots can help strengthen the defensive strategies of the ETI's security operations center by serving as a source of information for threat intelligence, technical staff training, and research into new attack trends. As future lines of work, it is proposed to integrate T-Pot with SIEM systems (such as OSSIM or ELK) to enhance automatic event correlation and to explore the deployment of distributed honeypots to expand the detection spectrum and increase network coverage.

In summary, the carried out research demonstrates that the implementation of honeypots constitutes a viable, scalable, and low-cost alternative for strengthening cybersecurity in critical infrastructures and Cuban enterprise environments.

References

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

RESEARCH

Proactive Training Against Phishing Emails Using Simulations with the GoPhish Tool

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

Received: 11/2025 | Accepted: 11/2025 | Published: 11/2025

Abstract

Phishing, a threat focused on the exploitation of the human factor, represents a high risk for organizations. Faced with the inadequacy of traditional theoretical training, this study implemented a controlled simulation program using the open-source tool GoPhish. Throughout four successive campaigns employing different attack vectors, a decreasing trend in interaction rates and a significant reduction in credential submission were observed. The analysis of participating users revealed a clear learning curve, where the majority showed substantial improvements in their detection capability. Concurrently, a considerable increase in voluntary reports of suspicious emails in the real environment was recorded, especially in high-exposure areas. The results prove that the combination of practical training, immediate feedback, and periodic simulations effectively transforms the user from a vulnerable link into a proactive defense barrier, thereby sustainably strengthening the organizational cybersecurity culture.

Keywords: phishing, GoPhish, cybersecurity simulation, practical training

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

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

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

Introduction

The current cybersecurity scenario is characterized by ever-evolving and increasingly sophisticated threats, with email remaining one of the most prevalent and effective attack vectors. Among these threats, phishing is defined as a form of online fraud whose objective is to steal confidential information by sending communications that impersonate the identity of a legitimate entity (Workman, 2008). This threat stands out not for its technical complexity, but for its ingenious exploitation of a vulnerable link: the human factor. Hackers have improved their tactics, creating fraudulent messages that are so convincing they can easily breach even the best-configured perimeter defenses by targeting employees' psychology directly.

The consequences of a successful phishing attack are severe, ranging from credential compromise and theft of sensitive information to ransomware infections that can paralyze a company's entire operations, generating substantial financial losses and irreparable damage to corporate reputation. In this situation, implementing only defensive technical solutions (such as email filters, antivirus software, and firewalls) has proven insufficient.

The most critical and underestimated defense lies in the continuous and conscious training of end users. To achieve this, companies must adopt a proactive approach that goes beyond traditional, theoretical training. This approach involves simulating controlled attacks, which allow companies to measure and improve their workforce's actual level of preparedness. In this context, phishing email simulation tools like GoPhish emerge as invaluable resources, allowing cybersecurity specialists to perform and analyze simulated phishing campaigns internally within companies. This practice not only identifies the most vulnerable employees to provide them with targeted training but also lays the foundations for a robust and resilient organizational cybersecurity culture, transforming the weakest link into an effective first line of defense.

The main objective of this paper is to measure the workforce susceptibility, evaluate the simulation tools effectiveness for

proactive training in detecting phishing emails, and identify areas for improvement.

Materials and Methods

Study Context and Approach Justification

The study was conducted in a logistics services company with more than 200 employees who use email daily, and more than 50 % maintain some type of exchange with international customers and suppliers, which exponentially increases interaction with real phishing emails; their identification knowledge is vital in detecting and minimizing incidents. To face this threat, a controlled simulation program was applied.

Phishing is not a new threat, but its prevalence and complexity continue to grow alarmingly. Annual reports from industry leaders provide frightening data proving the magnitude of the problem. The Verizon 2025 Data Breach Investigations Report (DBIR) consolidates this trend, placing this pattern among the top three since 2019 (Verizon, 2025). This report highlights that sophistication has evolved significantly; attacks are no longer based on incredible job offers or prizes, but on prolonged campaigns where hackers build trusting relationships with victims over a period of time.

The effectiveness of this technique lies in its social engineering, a method that "psychologically influences a person to perform an action or reveal confidential information" (ENISA, 2022). Hackers thoroughly research their victims, creating highly personalized emails (spear phishing) that perfectly mimic legitimate communications from superiors, colleagues, or trusted services. Verizon's DBIR (2025) validates that one must now be cautious even with messages that appear to come from established customers or suppliers, making detection very difficult even for experienced users.

The User, the "Weakest Link"

The traditional reference in cybersecurity has characterized the end user as "the weakest link" in the security chain. This perspective, however, is counterproductive and is being reevaluated. Research in the field of information security suggests that it is more effective

to see the employee not as a problem to be mitigated, but as a "human factor" that must be empowered and trained (Hadlington, 2017); "employees are not just users, but active guardians of cybersecurity" (Hernández, 2025). Users do not fail due to inherent negligence, but because the training they receive is often generic, sporadic, and disconnected from the actual tactics used by hackers today.

A study published in Computers & Security found that traditional awareness-based training has limited behavioral impact unless supplemented with hands-on experience and immediate feedback (Pattinson et al., 2015). This is where training with reality becomes crucial: zero trust is not only about limiting access, but about empowering people to take informed decisions (Hernández, 2025). By exposing users to realistic phishing simulations, experiential learning is activated, allowing them to recognize the signs of an attack (such as strange senders, suspicious URLs, or unjustified urgency) in a safe environment where a mistake becomes a lesson and learning, not a disaster.

Selected Tool: GoPhish

To implement a simulation-based training program, companies need accessible and effective tools that assess employee susceptibility through controlled attacks. Although the market is dominated by recognized commercial solutions, such as KnowBe4 and other leaders identified in the report "Forrester Wave: Security Awareness and Training Solutions, Q1 2022" (Budge, et al., 2022), platform choice often depends on budget and specific needs. For organizations that prioritize flexibility and the absence of licensing costs, GoPhish emerges as an open-source alternative with wide acceptance.

GoPhish is a platform that allows security professionals to:

- Design highly customizable phishing campaigns that replicate current tactics.
- Target these campaigns to specific groups of employees in a controlled manner.
- Accurately measure key metrics such as click-through rate and credential submission rate.

The value of GoPhish, backed by an active community and extensive documentation, lies in its ability to close the learning cycle: from theory to practice, and from practice to continuous improvement.

Campaign Design and Experimental Procedure

- Two different attack vectors were used for the simulations:
- Type 1: Sending an email requesting credential update on a web service.
- Type 2: Sending an email requesting file download.

Four simulated campaigns were performed over the course of two years. (2024-2025) (Table 1):

- Campaign 1 (Type 1): Credential update on the company's Webmail.
- Campaign 2 (Type 2): Request to download the file "Company Security Policy Update".
- Campaign 3 (Type 2): Request to download the file "Fair Product Listing".
- Campaign 4 (Type 1): Credential update on the corporate Enterprise Resource Planning (ERP).

Table 1. Campaigns in chronological order

	Users	Clic on this link	Submited credentials	Downloaded file	Users' report
Campaign 1	176	21	10	--	0
Campaign 2	173	17	--	17	8
Campaign3	227	40	--	40	8
Campaign 4	132	16	0	--	11

Corrective Measures and Data Collection

After each Campaign, and especially after the initial results, corrective actions were implemented. After Campaign 1, a specific Training Plan on phishing email identification was developed, aimed at those who did not pass the simulations, with detailed materials and practical identification exercises. Additionally, general training for all workers

on phishing emails and general awareness sessions with users were prepared, focused on the importance of proactively notifying any suspicious communication to the cybersecurity team, in order to speed up response to real incidents. Data was collected automatically by the GoPhish tool (clicks, credential submissions) and through manual user reports to the IT Management.

Results and Discussion

Results by Campaign and Evolution Analysis

Campaign 1 consisted of simulating an email that, through a fraudulent link, asked users to update their credentials (Figure 1) on a portal that emulated the corporate Webmail interface (Figure 2). After the user entered the data, the request was redirected to the company's official portal, recording the event as a successful phishing attempt. The results of this campaign revealed a 20% click-through rate and, of these, a 48% credential submission rate. These alarming initial figures evidenced a clear vulnerability of personnel to this type of attack, despite the existence of periodic training in the company. An additional critical finding was the almost non-existent notification of fraudulent emails by employees.

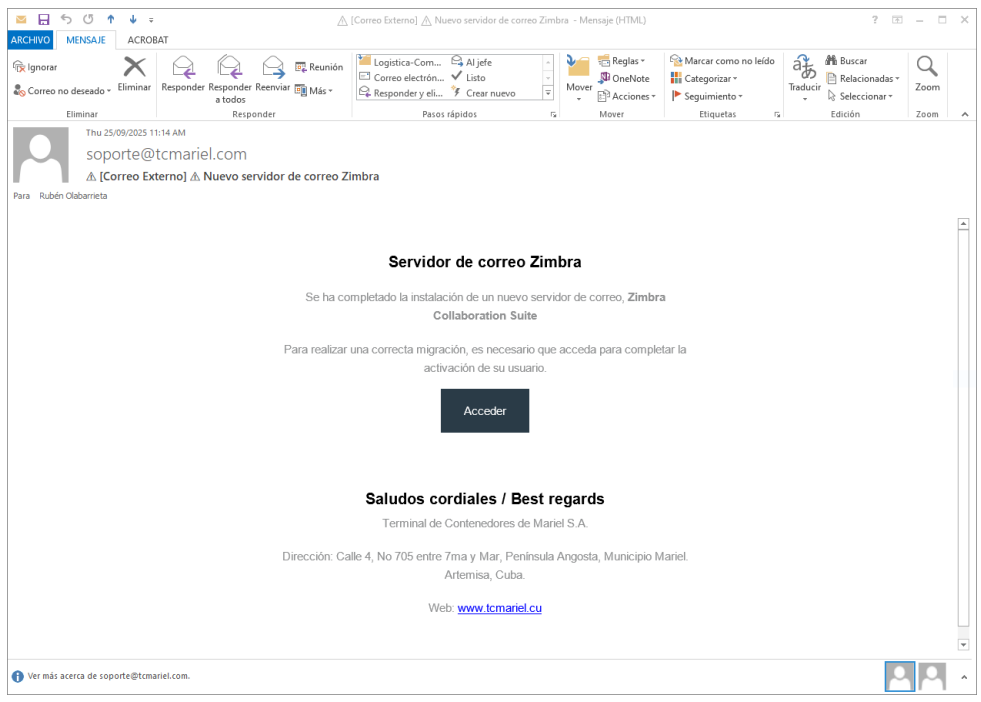


Figure 1. Email requesting credential update on the company's Webmail

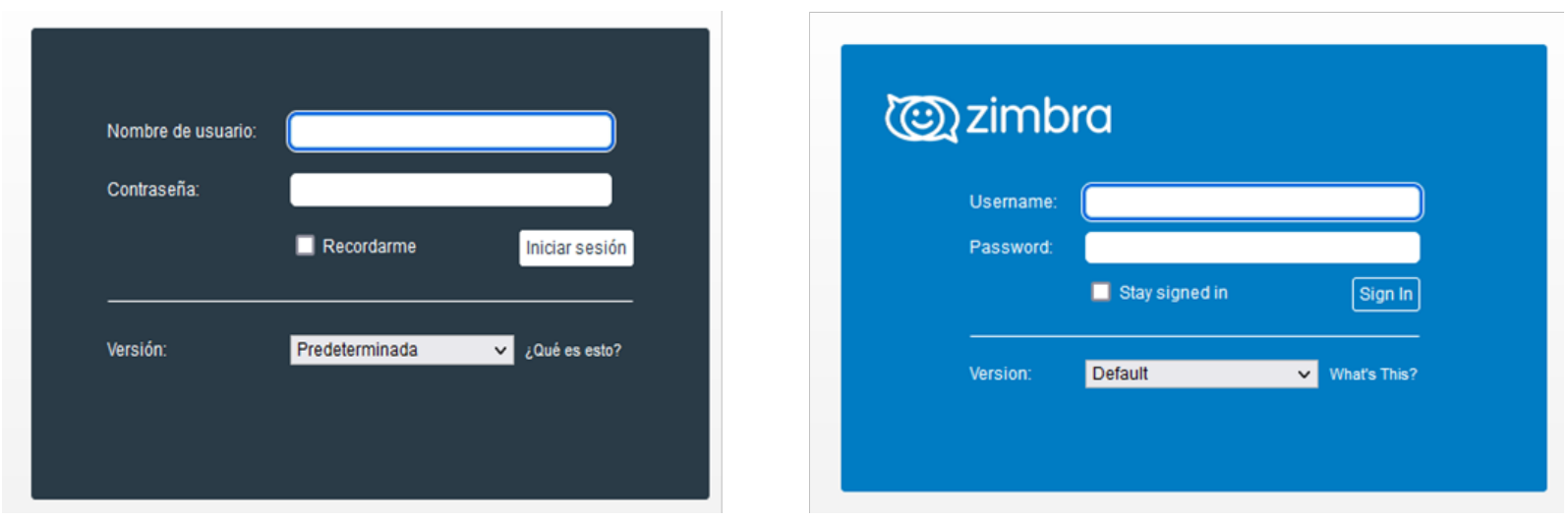


Figure 2. Fake (left) and real (right) corporate Webmail interface

Campaign 2 was designed to evaluate the effectiveness of the implemented awareness measures using a different attack vector: downloading a malicious file. In this simulation, the email urged users to download a file called "Company Security Policy Update". The event was recorded as a successful phishing attempt if the user proceeded to download the file. The results showed substantial improvement: the click-and-download rate dropped to 10%. Even more significantly, an increase in employee notifications was recorded, going from 0% in the initial campaign to 4%. This increase in reports, although modest, suggests greater awareness and a proactive response from personnel, demonstrating the preliminary effectiveness of the specific and general training provided after the first simulation.

Campaign 3 significantly increased the level of sophistication through the application of advanced social engineering techniques. The email, designed with greater adherence to the company's operational reality, requested the download of a file named "Fair Product Listing". Upon execution, this file initiated an automated sequence that simulated malicious behaviors, such as the consecutive execution of default Windows© system applications, culminating in the activation of a second script (Figure 3). This script displayed, on the workstations' desktop, through a command window (CMD), a graphical representation in the shape of a skull (Figure 4), aiming to more credibly emulate the observable consequences of a real attack.

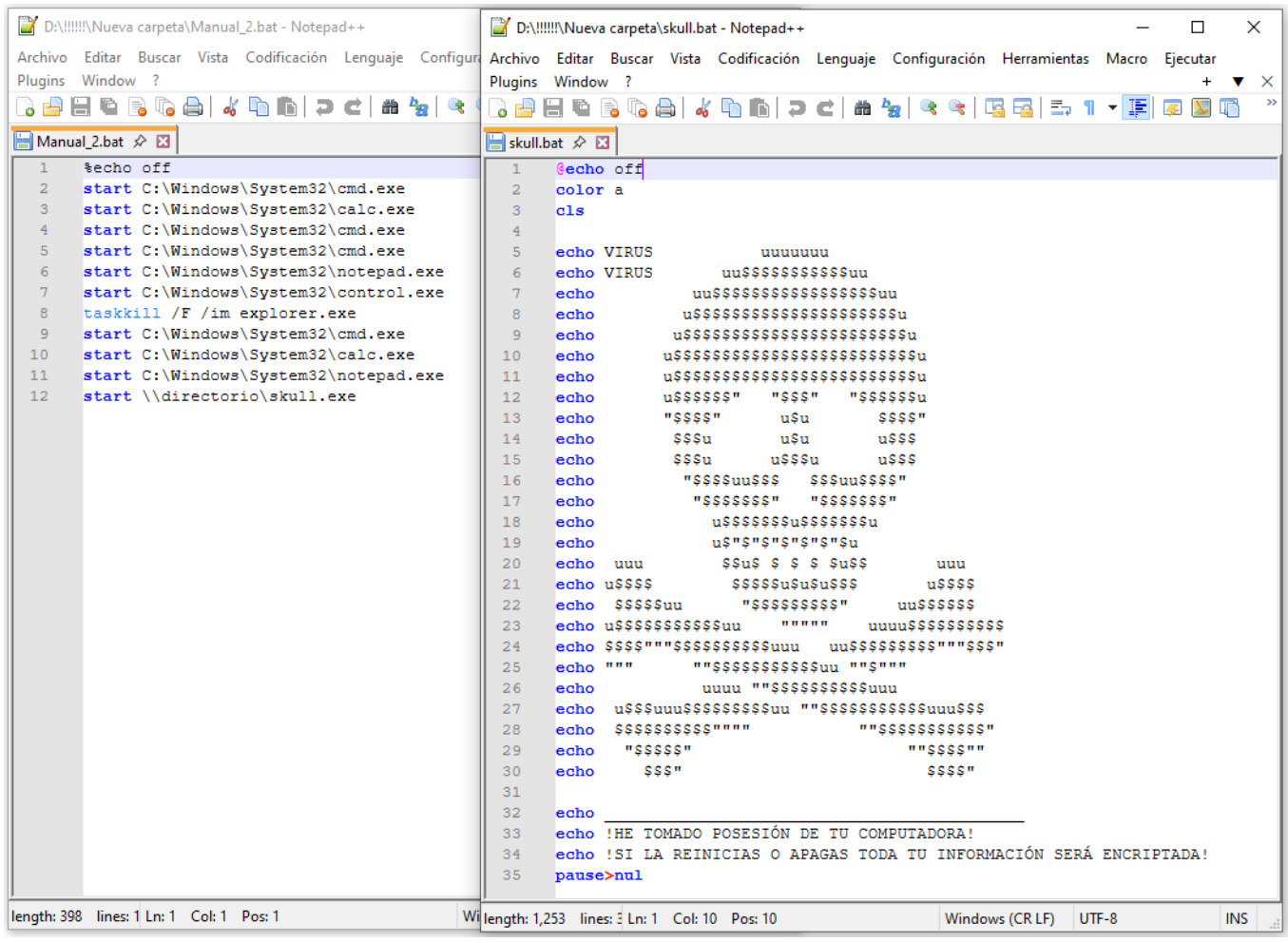


Figure 3. First (left) and second (right) script that executed when downloading the file

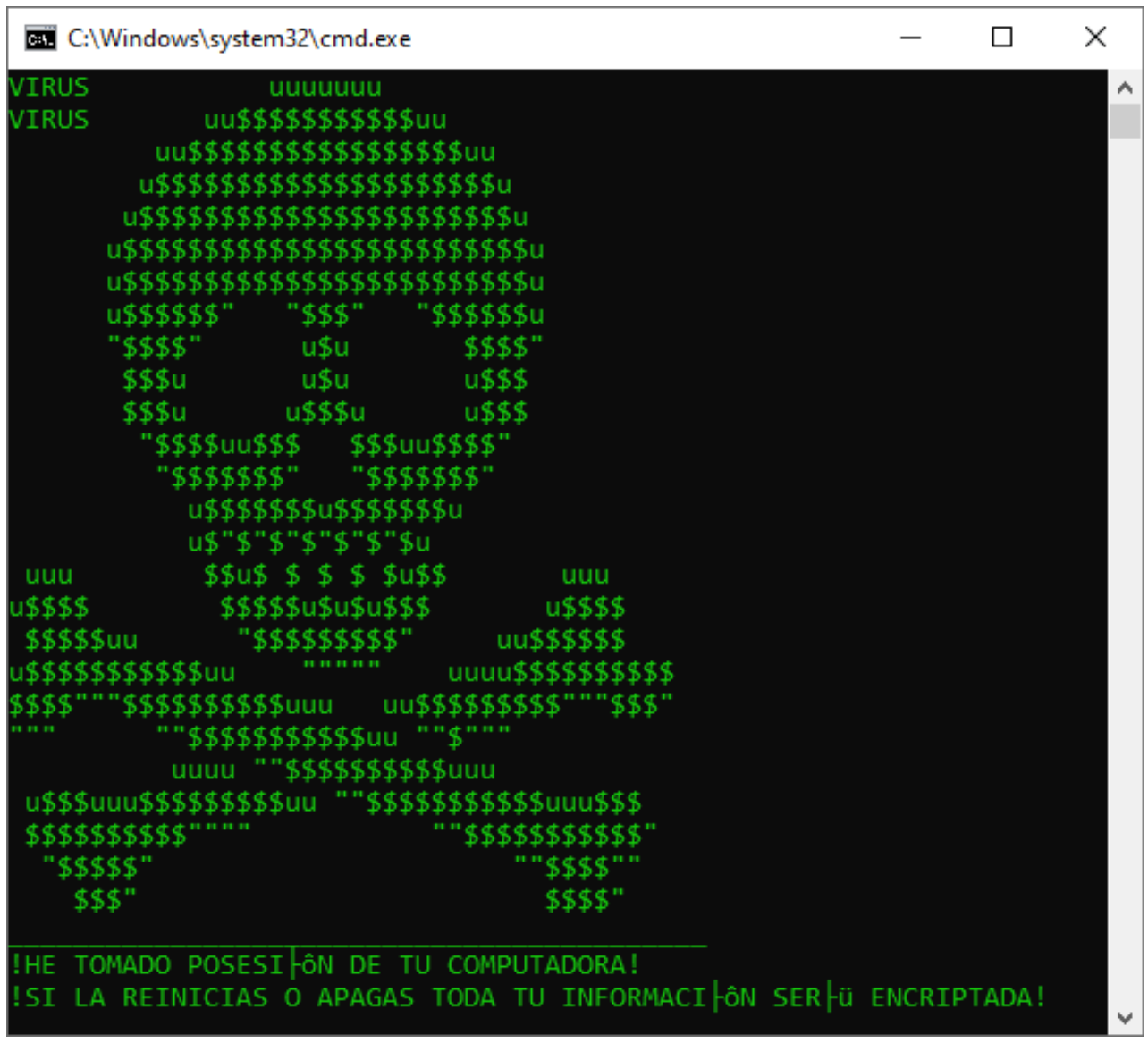


Figure 4. Graphical representation in the shape of a skull in a CMD window

The results of this campaign were revealing: although a considerable increase in interactions with the fraudulent email was recorded (Figure 5), the number of reports remained at eight users. Follow-up interviews conducted in the company's different departments identified the main cause of this apparent discrepancy. The notable visual impact of the previous simulation had generated internal conversations among workers, resulting in widespread knowledge of the test. Consequently, users who identified the email chose not to report it, recognizing it as an internal training element rather than a genuine external threat. This phenomenon underscores the importance of varying attack vectors in successive simulations to avoid result contamination and maintain the exercise's validity.

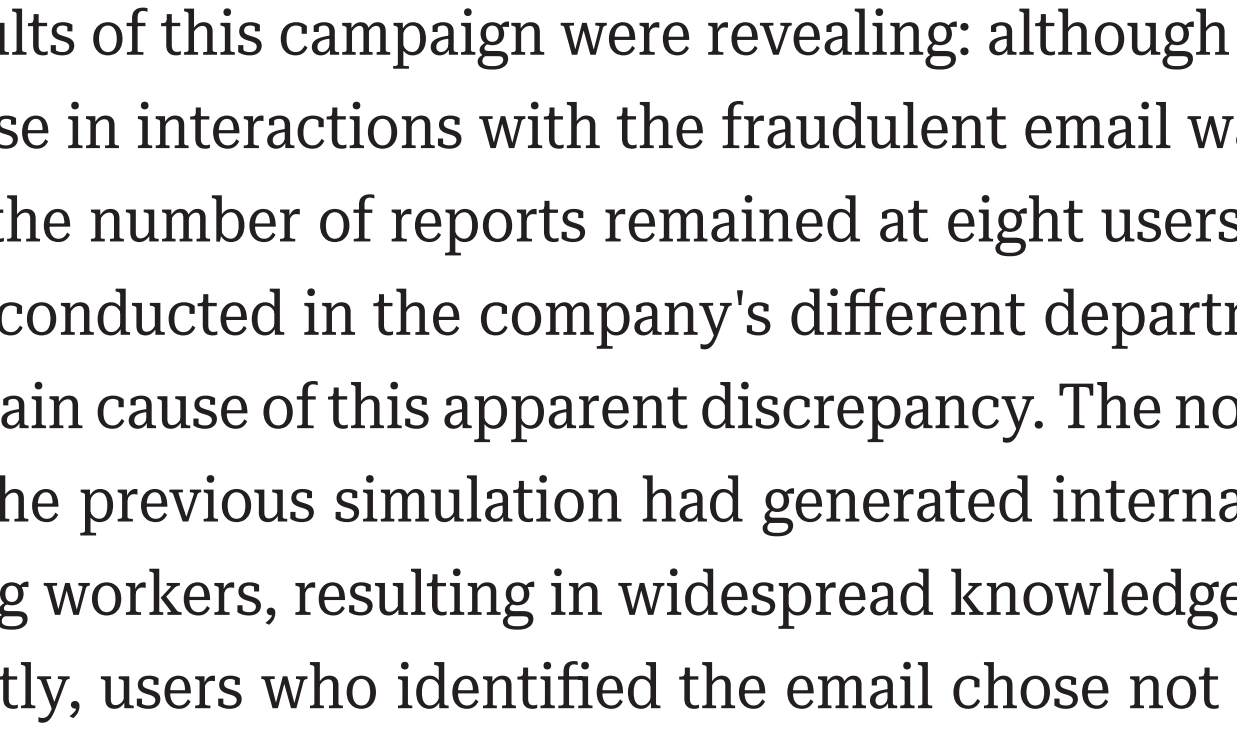


Figure 5. Number of interactions per Campaign

Campaign 4 returned to Campaign 1's attack vector (credential update request), but this time applied to accessing the corporate ERP system, with a redesigned phishing platform for this simulation. The results proved a significant improvement in user resilience. This campaign recorded the lowest level of interaction, with 16 users clicking on the fraudulent link. Even more notably, the credential submission rate was zero, as none of the users who accessed the fake page entered their data. At the same time, a considerable increase in the number of employee reports was observed (Figure 6). The comparative analysis of the four campaigns shows a clear trend of progressive improvement: the drastic reduction in interactions and phishing success rate, together with the sustained increase in notifications, confirm the effectiveness of the continuous training program.

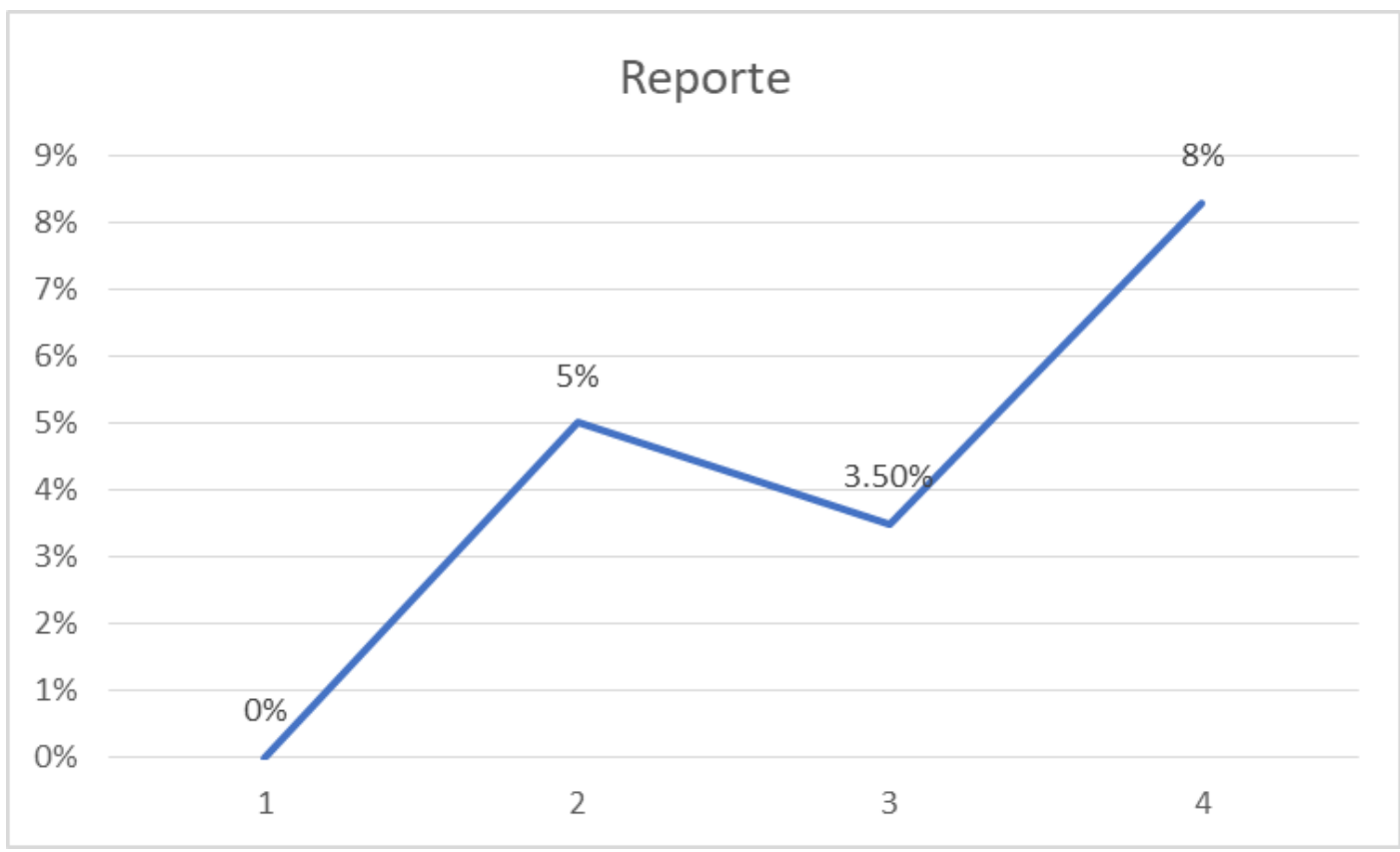


Figura 6. User reporting per Campaign

The decreasing trend in interaction and credential submission rates throughout the successive campaigns unequivocally shows the effectiveness of the implemented training program. A detailed analysis of the 227 users who participated in total across all simulations reveals that the majority showed a clear learning curve: while 46 users (20.3%) interacted in only one simulation, only 21 (9.3%) did so on two occasions, and only 2 users (0.9%) maintained vulnerable behavior in three of the four campaigns (Figure 7). This distribution shows that practical training and immediate feedback were effective for the vast majority of personnel.

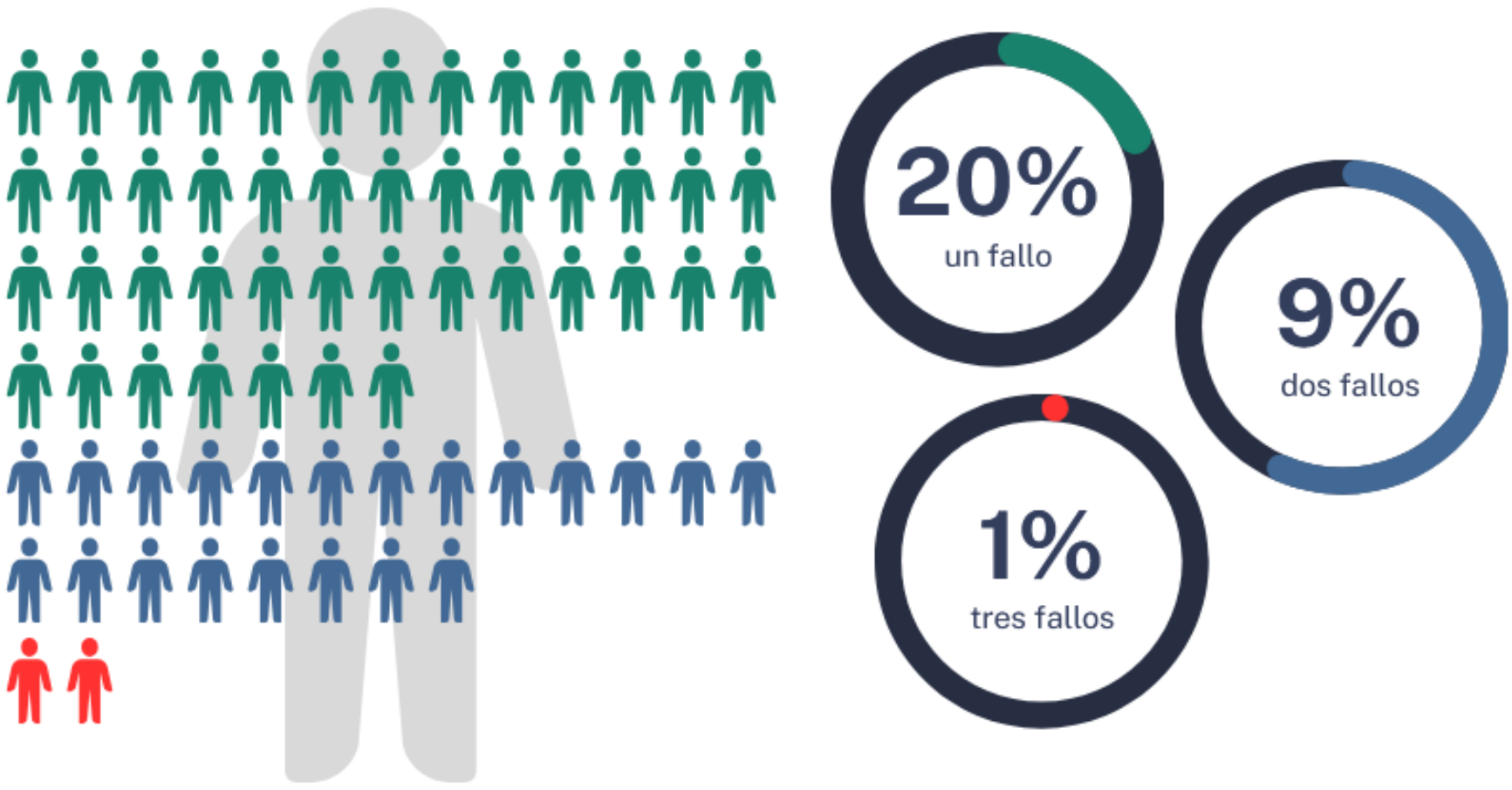


Figure 7. 7. Number of repeated failures by users

The success is reinforced by a key behavioral indicator: the IT Management reported a substantial increase in voluntary reports of suspicious emails by employees in the production environment, transcending the scope of simulations. A more detailed analysis reveals that the departments with the highest volume of real phishing email reports were Logistics and Commercial, which is consistent with their high exposure to international communications. This finding not only validates the effectiveness of the awareness efforts but also points to a palpable assimilation of security protocols by users and translates into a strengthening of the organizational cybersecurity culture. As a direct result of this increased vigilance, the effectiveness of proactive blocking of malicious domains has increased, thus closing a cycle of continuous improvement in the company's security posture.

Conclusions

The findings of this case study, in line with global statistics, prove that traditional theoretical training approaches are insufficient to effectively mitigate the risk associated with phishing.

Implementing simulation tools like GoPhish is a crucial strategy because it allows quantifying risk objectively by transforming abstract vulnerabilities into tangible metrics. This ability to present

concrete data (such as the specific percentage of employees who interact with a simulation) facilitates effective communication with senior management regarding the need to prioritize security resources.

Furthermore, these tools enable highly targeted training and resource optimization by accurately identifying the most vulnerable users. This approach allows designing personalized proactive awareness programs that maximize the impact of training initiatives. In parallel, continuous simulations foster a constant alertness state and a security culture, integrating cybersecurity into the daily routine and transforming human capital from a passive link into an active component of organizational defense.

The cost of implementing a solution like GoPhish is minimal compared to the potential cost of financial and reputational losses from a single successful data breach. Investing in the preparation of the human factor is, therefore, one of the most effective risk mitigation strategies available.

An organization's cyber strength can no longer depend solely on firewalls and software. The defense must be people-based as well. Training tools are fundamental to achieving the operability of this vision and facilitate companies transforming their workforce from a potential target into a robust and conscious defense barrier, thereby significantly and sustainably elevating their digital security posture.

References

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

RESEARCH

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

Laura Celín Sosa Martínez, ST*

Received: 11/2025 | Accepted: 11/2025 | Published: 11/2025

Abstract

This study addressed the implementation of a specialized tool for trace monitoring and analysis, 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 tool validation proved 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 is an effective strategy to ensure operational continuity and protect systems in public institutions.

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

* Directorate of Computer Security, University of Computer Sciences (UCI), Carretera a San Antonio de los Baños, Km 2 ½, reparto Torrens, municipio Boyeros, La Habana, Cuba.lsosa@uci.cu.

Introduction

Trace monitoring and analysis have become essential practices for the proper functioning of cumputer systems in public and private institutions. In the current context, where computer systems are increasingly interconnected and depend on technology to operate, it is essential to be able to monitor network and server activities so as to identify performance issues, process failures, and potential threats (Al-zahrani *et al*., 2025).

As organizations digitize and become more dependent on technological systems, information security and operational efficiency become key priorities. Traces record detailed information about each event within the systems, allowing not only precise diagnostics of problems, but also ensuring data confidentiality and integrity (Ahmed *et al*., 2023).

In the case of Cuba, trace monitoring and analysis have become a growing priority as the country advances in its process of the digital transformation of society. With the adoption of new technologies and the digitization of public services, the need to implement effective solutions to monitor computer systems is more important than ever. Despite technological and connectivity challenges, Cuba has made significant progress in strengthening its technological infrastructure, prioritizing the security and performance of its systems (Álvarez & Andrés, 2023).

The Prosecutor's Office located in Mayabeque Province, founded in 2010, aims to ensure compliance with the Constitution, laws, and other legal provisions, both by State agencies and citizens. Its mission is to exercise control over criminal investigations and public criminal prosecution on behalf of the State, ensuring strict regulatory compliance, and protecting data and information. This implies a commitment to the security and privacy of the information handled; all of the above guarantees not only transparency in its functions, but also respect for people's rights.

At the Mayabeque Provincial Prosecutor's Office, a serious deficiency regarding trace monitoring was evidenced. One of the most

worrying findings was the lack of an effective system to track and document information traces, which directly affects the Prosecutor's Office's ability to detect anomalies and recurring failures in its computer systems, thus jeopardizing the efficiency of its services.

Based on the abovementioned explanation, the following research problem arises: How to contribute to the improvement of the monitoring and analysis of traces at theg Mayabeque Provincial Prosecutor's Office?

To solve this problem, the overall objective is to implement a trace monitoring and analysis tool at the Mayabeque Provincial Prosecutor's Office in order to detect anomalies.

Materials and methods

To select the appropriate tool for trace monitoring and analysis, an exhaustive search was conducted for technological solutions recognized for their advanced analysis capacity and effectiveness in detecting problems. Among the options evaluated, Grafana, Prometheus, and the ELK Stack were identified, all open-source and widely used in observability and system supervision environments.

Grafana is a visualization tool that analyzes data in real-time through interactive and customizable dashboards. It integrates with multiple sources such as Prometheus, Elasticsearch, and InfluxDB and offers functionalities like configurable alerts, OpenTelemetry integration, and graphical dashboard design (Leppänen, 2021).

Prometheus, on the other hand, is oriented towards collecting metrics in the form of time series. Its multidimensional data model and its PromQL query language allow detailed analysis and generation of custom alerts. Although its visualization capabilities are limited, it effectively complements Grafana (Kim et al., 2024).

The ELK Stack stands out for its ability to centralize and analyze traces from multiple sources in distributed systems. Its scalable architecture manages large volumes of data, correlates events in real-time, and detects behavioral patterns. Additionally, it facilitates detailed visualization through Kibana, improving data interpretation and decision-making (Gatsi *et al.*, 2023).

To meet the specific needs of Mayabeque Provincial Prosecutor's Office, key criteria for tool selection were defined: compatibility with existing systems, zero cost due to being open-source, advanced trace analysis capability, data search efficiency, ease of use for non-technical staff, and extensive visualization options.

The following is a comparative table (Table 1) summarizing the characteristics of the evaluated tools:

Table 1. Comparison of tools for trace monitoring

Tool	Compatibility	Cost	Trace Analysis	Data Search	Usability	Visualization
ELK Stack	Integrates with Windows, Linux, MacOS	Free	High capacity to detect patterns	Fast and efficient searches	Intuitive and flexible interface	Dynamic and real-time visualization
Prometheus	Integrates with Windows, Linux, MacOS	Free	Good capacity for complex patterns	Advanced queries	Requires technical knowledge	Limited visualization, complements with Grafana
Grafana	Integrates with Windows, Linux, MacOS	Free	Limited, depends on external integration	Does not perform direct searches	Customizable graphical interface	Customizable graphical interface

Based on this comparison, it was determined that the ELK Stack is the most suitable tool to put in practice at the Mayabeque Provincial Prosecutor's Office. Its versatility, analysis capacity, data search efficiency, and ease of use make it the ideal solution to improve trace monitoring.

Kibana's graphical interface allows for rapid adoption by staff, while integration with Logstash and Elasticsearch ensures effective log centralization. The ELK Stack meets all established requirements, offering a robust platform for system supervision, anomaly detection, and generation of visual reports that facilitate informed decision-making.

Results and discussion

The ELK Stack tool was implemented in a controlled environment within the Mayabeque Provincial Prosecutor's Office to validate its effectiveness in trace monitoring and analysis. The main components were configured: Logstash for data collection and processing; Elasticsearch for efficient storage and searching; and Kibana for interactive log visualization.

During its implementation, specific environment variables were defined (Figure 1) that allowed the tool to be adapted to the characteristics of the Prosecutor's Office's technological infrastructure. The

functioning of the indices generated by Elasticsearch was verified, and the Kibana server was activated for real-time visualization of registered events.

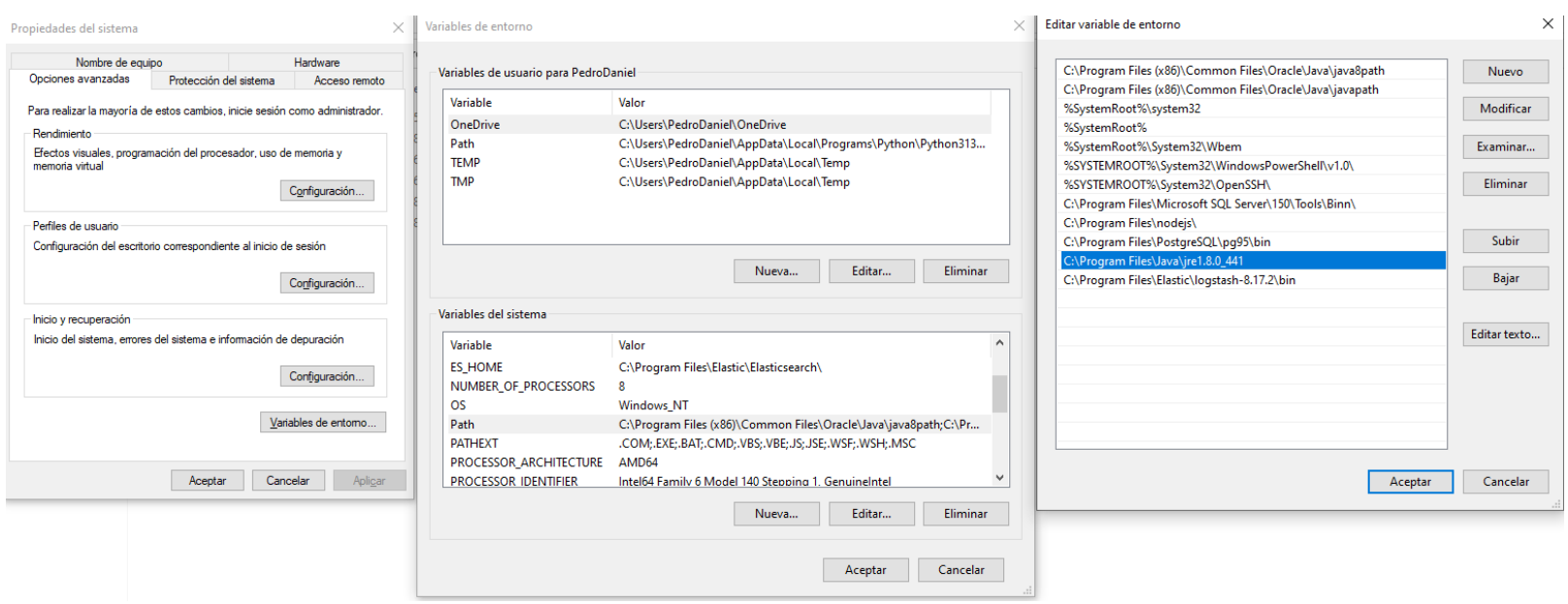


Figure 1. Environment variables

Once configuration was completed, functional validation of the tool proceeded. Tests were conducted for event detection, alert verification, and trace analysis in various simulated scenarios (Figure 2). The results showed that the tool was capable of identifying anomalous patterns, classifying threats according to their impact level, and generating timely alerts for potential failures (Figure 3 and Figure 4).

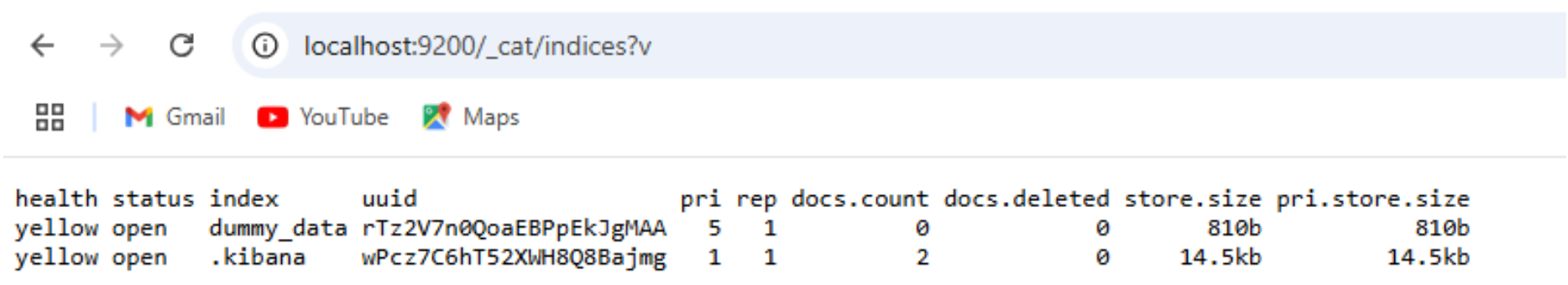


Figure 2. Index verification

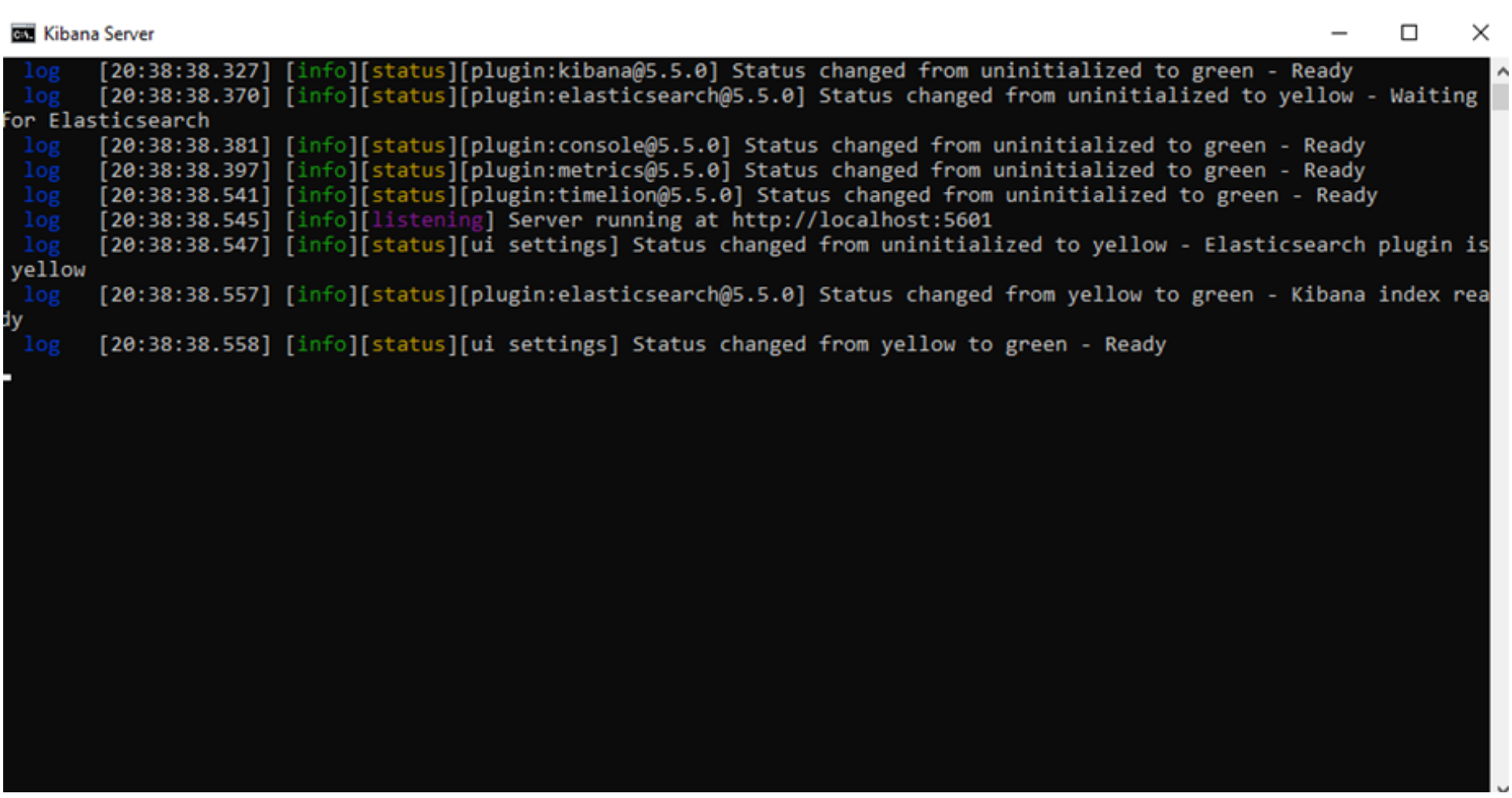


Figure 3. Active Kibana server

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 }
}
```

Figure 4. Logstash configuration

The visualization of detected events allowed technical staff to interpret data clearly and quickly, facilitating informed decision-making. Additionally, a significant improvement in incident response capability was achieved, contributing to optimized supervision of information systems (Figure 5; Figure 6; Table 2; Table 3; Table 4).

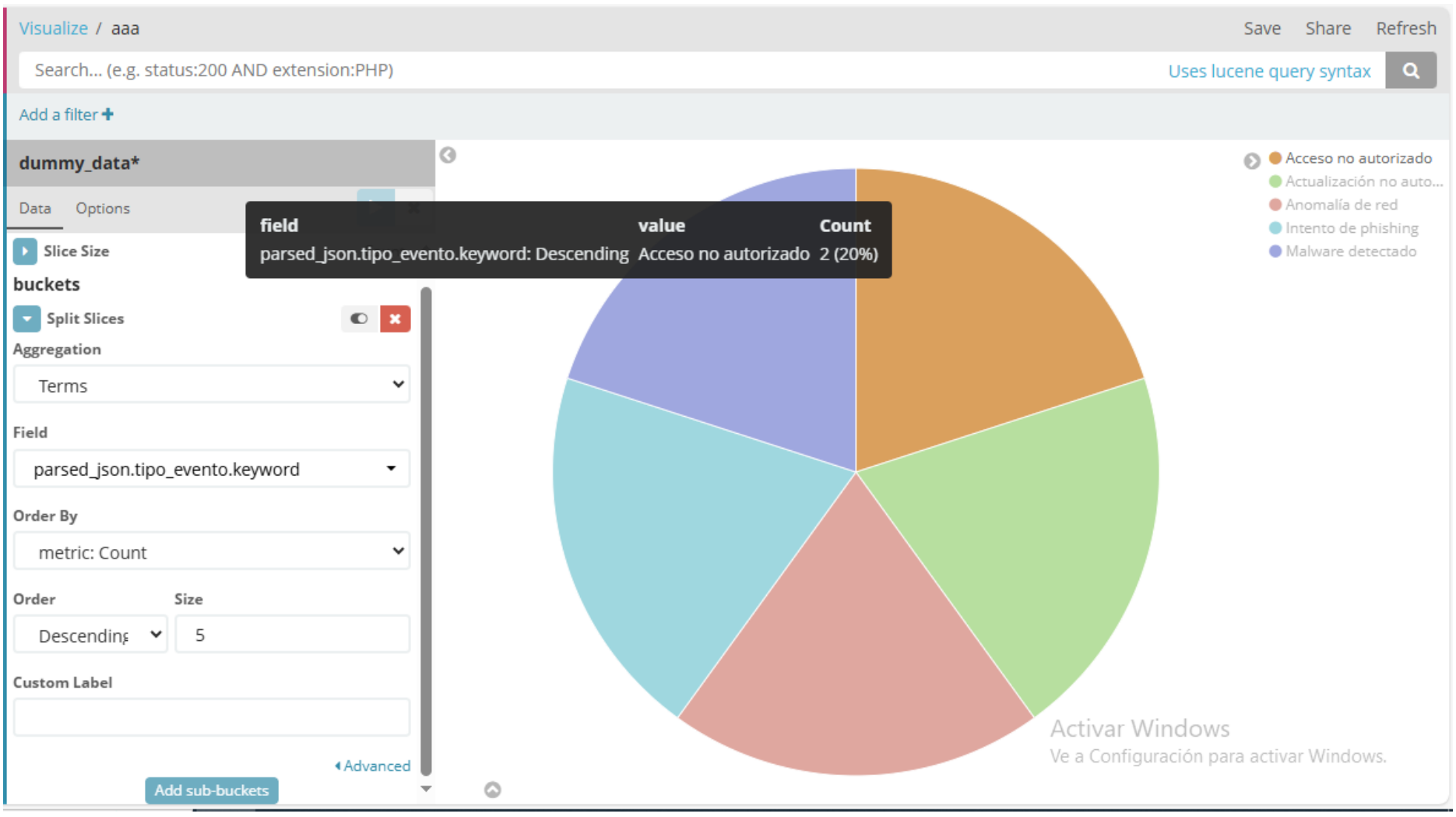


Figure 5. Detected events

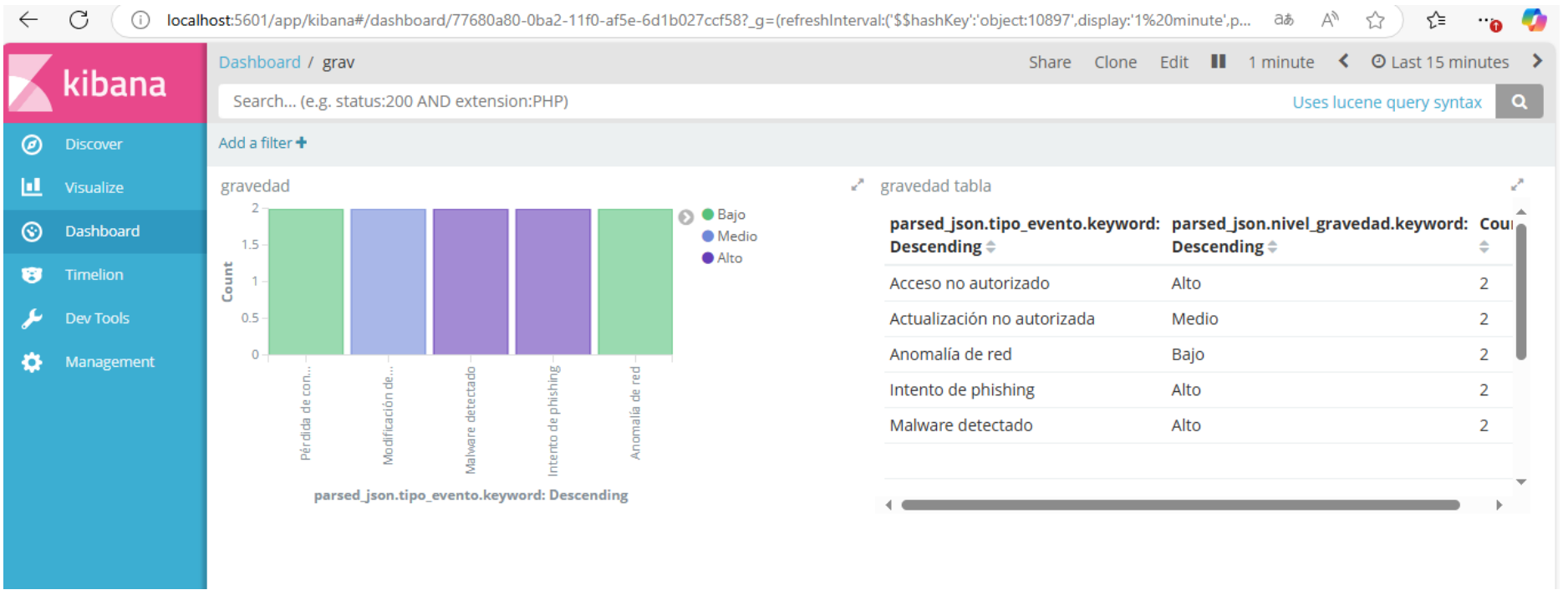


Figure 6. Threat classification

Table 2. Classification of detected events

Event	Quantity Detected	Severity	Action Taken
Unauthorized access attempts	15	Critical	IP blocking and firewall adjustments
Suspicious downloads	10	Medium	Download restrictions
Suspicious activities	5	Low	Continuous monitoring

Table 3. Classification of identified threats

Identified Threat	Description	Severity Classification	Recommended Actions
Exposed vulnerable service	Deficient configuration allowing unauthorized access from the Internet	Critical	Update configurations and apply strict policies
Unauthorized downloads	Downloads that can introduce malware into the system without user intervention	Medium	Restrict downloads
Social engineering-based attacks	Malicious applications delivered via email or downloaded by users	Low	Implement email filters, educate users

Table 4. Summary of problems and resolutions

Problem	Corrective Action	Status
Exposed vulnerable service	Reconfiguration and application of restrictions	Resolved
Suspicious activities	User training and continuous monitoring	In process

The results obtained show that the proposed solution constitutes a substantial improvement over previous practices. Log centralization, advanced trace analysis, and dynamic visualization represent a novelty in the context of the Mayabeque Provincial Prosecutor's Office, where there was not tool with these capabilities previously.

The ELK Stack proved to be superior to other similar solutions in terms of functionality, adaptability, and ease of use, confirming the fulfillment of the objectives set in this paper. Its implementation not only strengthened the institution's IT security but also laid the basis for future improvements in technology management.

Conclusions

The implementation of the ELK Stack tool at the Mayabeque Provincial Prosecutor's Office significantly improved the monitoring and analysis of traces in its computer systems. This technological solution effectively responded to the problem identified in the audit by providing a platform capable of centralizing logs, detecting relevant events, and visualizing data clearly and dynamically.

The general objective of the study was achieved by establishing a tool that facilitates anomaly identification, strengthens the security of the technological infrastructure, and optimizes operational supervision. The results obtained prove that the use of traces with visual support represents an efficient strategy to ensure the continuity of institutional services.

The work carried out constitutes a relevant contribution in the field of network administration and IT security, showing that it is possible to implement advanced solutions with accessible, open-source resources. The experience gained in this process can serve as a reference for other public institutions facing similar challenges in managing their technological systems.

As future work, it is recommended to continue developing the tool by incorporating artificial intelligence mechanisms for predictive threat detection, as well as expanding its integration with other existing monitoring systems. Furthermore, it is suggested to train technical staff regarding the use of these technologies so as to ensure their full and sustainable utilization.

References

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.

RESEARCH

Implementing artificial intelligence models for malware detection in critical telecommunications infrastructures

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

Received: 11/2025 | Accepted: 11/2025 | Published: 11/2025

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 to key strategy for ensuring the resilience, availability, and continuity of essential services. This paper 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

[*] Engineer. Department of Computer Security, University of Computer Sciences. elizabethmm@uci.cu

[2] Engineer. Department of Technology Infrastructure, Faculty of Cybersecurity, University of Computer Sciences. heidyrm@uci.cu

[3] Master Degree. Department of Computer Security, University of Computer Sciences. henryraul@uci.cu

methodology focuses on to systematic review of recent documentation and to comparative analysis of models based on performance metrics, generalization capability, and computational resource consumption. Results indicate that hybrid models combining deep learning and network flow analysis techniques can achieve detection rates of over 98% with a significant reduction in false positives. 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

Introduction

Critical telecommunications infrastructure are the backbone of modern communication systems, providing essential services for the functioning of governments, businesses, and citizens. This infrastructure includes transmission networks, data centers, switching systems, and management platforms that ensure national and international connectivity. However, their growing complexity and reliance on software make them a prime target for cyberattackers (García-Teodoro *et al*., 2023).

Malware (a term encompassing viruses, Trojans, ransomware, spyware, and rootkits) has evolved in sophistication, employing evasion techniques such as code obfuscation, polymorphism, and "living-off-the-land" (LotL); this hinders detection by traditional signature-based mechanisms (Alasmary et al., 2022). In response to this challenge, artificial intelligence (AI) has emerged as a fundamental tool for the automated identification of anomalous patterns and the real-time detection of emerging threats (Sgandurra & Lupu, 2020).

Machine Learning (ML) and Deep Learning (DL) have proved remarkable potential for classifying malware from large datasets by extracting both static features (from binary code) and dynamic features (from runtime behavior). These models learn complex

representations that identify new or unknown variants, even in scenarios involving targeted attacks against critical infrastructure (Ucci, Aniello & Baldoni, 2019).

This paper aims to analyze the application of artificial intelligence models for malware detection within strategic telecommunications sectors. The issue is addressed from a technical-scientific perspective, combining the study of algorithms, feature extraction techniques, and the operational context of telecommunications networks. Furthermore, the study examines challenges related to model interpretability, the management of large datasets, and integration with existing security systems, such as Security Information and Event Management (SIEM) and Intrusion Detection Systems (IDS).

This paper is structured as follows: the Materials and Methods section describes the methodology used for the systematic review and classification of AI models; the Results and Discussion section analyzes the key findings and their applicability in real-world environments; and the Conclusions section summarizes the contributions presented and proposes future lines of research aimed at strengthening cyber-resilience in the telecommunications sector.

Materials and Methods

For the elaboration of this study about the application of Artificial intelligence models in malware detection in critical telecommunications infrastructures, a methodology of systematic review and comparative analysis of existing approaches was adopted, complemented with a characterization of the most representative models implemented in telecommunications real environments.

Sources of Information and Selection Criteria

The documentation review was conducted in scientific databases of high impact such as IEEE Xplore, SpringerLink, ScienceDirect, ACM Digital Library, and Scopus, using combinations of words clue as "AI-based malware detection," "deep learning for cyber defense", "critical infrastructure protection," and "telecommunication cybersecurity."

The following inclusion criteria were established:

- Peer-reviewed publications between 2018 and 2025.
- Studies focused on detection or classification of malware using AI.
- Research applied to critical infrastructures or networks telecommunications.
- Jobs that reported quantitative performance metrics (precision, recall, F1-score or false positives rate).

Additionally, studies with unverifiable datasets or models lacking documented reproducibility were excluded. In total, 96 scientific publications were reviewed; of these, 38 met the established criteria and were included in the primary analysis.

Analysis Methodology

The methodology adopted was structured in three phases:

1. Systematic documentation review: The PRISMA protocol (Preferred Reporting Items for Systematic Reviews and Meta - Analyses) was used to identify and filter relevant studies. The initial search yielded 327 results, which were reduced after the implementation of the filters of quality, year and scope of application.

2. AI model classification: the models were grouped in four main categories according to their algorithmic nature:

- Classic ML models: SVM, Decision Trees, Random Forest, Naïve Bayes, KNN.
- DL models: CNN, RNN, LSTM, Autoencoders, GNN.
- Hybrids models: combinations of ML and DL, as CNN+SVM or Autoencoder+Random Forest.
- Models of detection based on network flow and performance analysis, especially in 5G and SDN settings (Software Defined Networking).

3. Comparative assessment: an analysis of the metrics of performance of the models was carried out, comparing accuracy, sensitivity and inference time. A weighting scheme based on the geometric mean of the precision and recall metrics was used to determine the relative efficiency of each model (Zhang et al., 2022).

Critical Infrastructures and Application Context

In the field of telecommunications, the critical infrastructures include fiber optic backbone networks, systems of signaling SS7 and SIP, nodes of commutation, base stations, backbone routers and network functions virtualization platforms (NFV). These infrastructures support key services like telephony, data, radio communication and satellite networks (ENISA, 2023).

Due to its strategic character and cross-sector dependence, a malware attack can have cascading effects on energy, financial and governmental services (CISA, 2022). Therefore, early detection of anomalies using AI is considered an essential component of operational cyber resilience (Shafiq et al., 2020).

Machine Learning Models Being Used

The classic algorithms of supervised learning such as Support Vector Machine (SVM) and Random Forest (RF) have shown a good performance in malware detection through static feature extraction (Permana *et al*., 2020).

SVM stands out by its ability to separate classes with optimal margins, while RF offers sturdiness in view of over adjustment to the combine multiple trees of decision.

However, deep learning models have surpassed these results in dynamic behavior detection scenarios. Convolutional Neural Networks (CNNs) have been successfully applied to byte sequence analysis and the identification of malware binary patterns (Kolosnjaji *et al*., 2018), while Long Short-Term Memory (LSTM) models excel at detecting malicious patterns in real time in network flows (Pektaş & Acarman, 2020).

Meanwhile, Autoencoders and Graph Neural Networks (GNNs) are used for unsupervised anomaly detection, learn high-dimensional representations of normal traffic, and detect significant deviations (Wang *et al*., 2022).

Data and Experimental Environments

The most frequently used databases in the reviewed studies were:

- EMBER Dataset (Endgame Malware Benchmark for Research).
- CICIDS2017 and CSE-CIC-IDS2018 (network traffic datasets)

- VirusShare and Microsoft Malware Classification Challenge (BIG 2015) for static analysis.
- NSL-KDD and TON_IoT datasets, relevant in IoT and 5G networks.

Implementation and Tools Used

The review showed that the most commonly used environments for experimentation are Python with TensorFlow, Keras, PyTorch, and Scikit -learn, along with cloud orchestration platforms such as Google Collab, AWS SageMaker and Azure ML. In research focused on critical environments, the models were deployed on SDN networks through mininet and OpenDaylight, with support for real time analysis (Al-Yaseen *et al*., 2020).

Limitations of the Study

Although the analysis offers a global overview, some limitations are identified:

- Scarcity of studies focused specifically on backbone networks and signaling systems.
- Lack of availability of specific critical telecommunications datasets.
- Difficulty reproducing certain results due to data confidentiality.
- Lack of explainable models that allow for the interpretation of algorithmic decisions in the face of complex incidents.

Results and Discussion

The results of the systematic review and comparative analysis of AI models applied to malware detection in critical telecommunications infrastructure prove the sustained evolution of detection methods toward approaches based on deep learning and adaptive contextual detection. This section presents the key findings, organized by model type, performance, metrics used, and practical implications within the context of telecommunications networks.

1. Performance of Machine Learning Models

Classical machine learning models (SVM, Random Forest, Naïve Bayes, Decision Tree) remain competitive in scenarios where malware detection relies on static features extracted from executable files or

the analysis of traffic headers. According to Permana *et al.* (2020), Random Forest achieved 95.6% accuracy in malware classification, while SVM delivered similar performance with faster inference times.

However, their detection capability diminishes when faced with polymorphic or metamorphic malware samples, as well as in scenarios involving encrypted traffic or advanced evasion techniques. In the context of critical telecommunications networks—characterized by heterogeneous, high-speed traffic flows—classical models tend to generate high false-positive rates (FPR > 6%), which can overwhelm alert systems in monitoring centers (Cruz **et al.**, 2021).

Consequently, organizations in the sector have begun adopting hybrid models that combine the interpretability of classical methods with the generalization power of deep learning.

2. Effectiveness of Deep Learning Models

The analysis revealed that models based on deep neural networks (DNN, CNN, RNN, LSTM) offer substantially superior performance in detecting dynamic malware and classifying complex network traffic patterns. In particular, CNNs have been widely used to represent malware binary code as a matrix image and detect hidden malicious patterns.

Kolosnjaji et al. (2018) showed that a CNN trained on binary images of executables achieved 98.4% accuracy in detecting malware families without needing to disassemble the code. Meanwhile, LSTM and GRU are effective in detecting persistent malicious behavior by capturing temporal relationships between events in traffic sequences (Pektaş & Acarman, 2020).

Autoencoder models stand out in the detection of anomalies by reconstructing features. By being trained exclusively with legitimate traffic, they manage to identify deviations with accuracy rates above 97 %, which significantly reduces false positives in high availability environments (Wang et al., 2022).

On the other hand, GNNs and Transformers are emerging as benchmark models, capable of analyzing structural dependencies among network nodes, identifying coordinated attack flows, and correlating alerts across distributed domains (Hussain et al., 2023).

Table 1 summarizes the average metrics obtained from the main studies reviewed, comparing the models according to accuracy, recall, and false positive rate (FPR).

Table 1. Comparative Performance of AI Models in Malware Detection

Model Type	Algorithm example	Accuracy (%)	Recall (%)	FPR (%)
Classic ML	Random Forest	95.6	94.2	6.1
Classic ML	SVM	94.9	92.8	5.8
DL – CNN	CNN 2D in binary	98.4	97.2	2.3
DL – LSTM	Traffic sequences	97.6	96.9	3.1
DL – Autoencoder	Legitimate traffic	97.2	95.8	2.6
GNN / Transformer	Network topology	98.9	98.1	1.9

3. Applicability in Telecommunications Infrastructures

In telecommunications networks, the data flows are characterized by their high dimensionality, diversity of protocols and low latency, which makes it difficult to detect anomalous behavior in real time. AI models must adapt to this context by techniques of reduction of dimensionality and optimization of edge inference.

Several studies (Shafiq et al., 2020; Al-Yaseen *et al.*, 2020) have implemented CNN-LSTM models integrated into Software Defined Networking (SDN) environments and have detected botnet, ransomware, and denial-of-service attacks with a response time of less than 200 ms. In experimental scenarios simulating 5G telecommunications infrastructures, the models demonstrated a 35% reduction in detection latency and a 40% improvement in readiness compared to conventional methods.

Furthermore, the combination of Autoencoder models with flow analysis-based detection systems made it possible to isolate suspicious behavior in NFV (Network Function Virtualization) virtualization environments, ensuring operational continuity and proactive threat mitigation (ENISA, 2023).

In the case of telecommunications-related IoT networks, it was shown that GNN models are particularly effective in representing topology as a dynamic graph, identifying relationships between committed devices and coordinated events (Zhang et al., 2022).

4. Advantages and Challenges of Applying AI

The main advantages of AI models in malware detection include:

- High precision in the known and unknown threats IDs.
- Adaptability in view of changes in the attack patterns.
- Ability to continuously learn by updating parameters and datasets.
- Reduction of false positives, optimizing the resources of the security operations centers (SOC).

However, the challenges are equally significant:

- Most DL models are black boxes that are difficult to interpret (Goodman & Flaxman, 2017).
- Dependence on labeled data: the scarcity of telecommunications-specific datasets limits the training of specialized models.
- Vulnerability to adversarial attacks, which can deceive AI models by slightly modifying input patterns (Huang *et al*., 2021).
- High computational requirements, which hinder implementation in resource-constrained real-time systems.

5. Integration Into Telecommunications Security Systems

The deployment of AI models for malware detection has been progressively integrated into existing security architectures, such as Security Information and Event Management (SIEM) systems and AI-based Intrusion Detection Systems (IDS). In critical telecommunications environments, this integration requires intelligent orchestration capable of correlating events from multiple layers (physical, logical, and application).

Systems such as IBM QRadar, Splunk, and Cisco SecureX have begun incorporating deep learning-based detection modules to improve alert correlation in 5G networks (CISA, 2022). These hybrid approaches enable real-time detection and automated response, reducing the mean time to detect (MTTD) by more than 60%.

Furthermore, federated AI frameworks are being developed to enable collaborative model training among telecommunications operators without sharing sensitive data, thereby ensuring privacy

through federated learning techniques (Yang *et al.*, 2019). This approach holds promise for collective defense against large-scale cyberattacks.

Conclusions

The study proves that AI is an essential tool for strengthening the security of critical telecommunications infrastructure, offering malware detection solutions that are more accurate, adaptive, and scalable than traditional methods.

Comparative results indicate that deep learning-based models—specifically CNNs, LSTMs, and GNNs—significantly outperform classical machine learning approaches in terms of accuracy, false-positive rates, and the ability to detect unknown threats. Hybrid models combining static, dynamic, and network flow analysis have yielded outstanding results in high-criticality environments, achieving detection rates exceeding 98%.

Furthermore, integrating AI with existing security systems, such as IDS and SIEM, enables automated event correlation and real-time response, thereby enhancing the capacity to detect and mitigate complex cyberattacks.

However, relevant challenges remain. These include the limited availability of data representative of critical infrastructures, the limited interpretability of deep learning models and vulnerability to adversarial attacks, which can manipulate predictions through minimal disturbances in the input data. Additionally, the high computational demands of some models limit their deployment on resource-constrained systems or at network edge nodes.

Consequently, it is recommended to advance the development of explainable AI (XAI) models that allow for an understanding of algorithmic decisions, the creation of datasets specific to the telecommunications sector, and the adoption of federated approaches that facilitate collaborative training among operators without compromising data privacy.

The application of AI to malware detection represents not merely a technical improvement but a strategic step toward comprehensive

cyber-resilience in telecommunications, ensuring the operational continuity of services that underpin the digital economy and national security.

References

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

RESEARCH

Integrated platform for digital signatures within the context of XETID Company

Huber Martínez Rodríguez, PhD[*], and Yosvel Dupeirón Porra, Eng.[2]

Received: 11/2025 | Accepted: 11/2025 | Published: 11/2025

Abstract

In the current process of digitizing of 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. The digital signature ensures the integrity and the signer's association with the signed digital document, thereby ensuring its authenticity for legal purposes. This paper proposes a platform, which integrates different components, with the aim of providing a secure digital signature service from the Information Technology Company for Defense (XETID, for its acronym in Spanish) for Cuban society in general. This solution involves a public key infrastructure, a digital signature server, a timestamp server, and a secure key and cryptographic operation management service. Additionally, tool typologies for utilizing the digital signature service are proposed.

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

[*] Information Technology Company for Defense, XETID, La Habana, Cuba. huber@xetid.cu

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

[2] Information Technology Company for Defense, XETID, La Habana, Cuba. ydupeiron@xetid.cu

Introduction

The digital signature has become an increasingly demanded tool by institutions and Cuban society in general. In the service digitalization process, the digital signature is crucial because it guarantees the integrity and association with the signer of a digital document, thereby ensuring its authenticity and legal validity. But it is not only about validating documents, the digital signature is the key that allows reducing papers, streamlining processes and guaranteeing legal traceability in each document.

In 2016, Cuba began the validation of digital signatures through the adoption of the public key infrastructure (PKI), in accordance with the guidelines set out by Resolution 2 of the Ministry of Internal Affairs (MININT, for its acronym in Spanish), later repealed by Decree-Law 79/2023 and, its regulation, Decree 106/2024 . This infrastructure ensures the issuance of digital certificates, which are electronic documents that verify the identity of a person or entity; as such, they can be associated with a pair of keys (**a public key and a private key**) that allow users to establish their identity in the digital world.

In 2018, Decree-Law 370 on the informatization of Cuban society, legally recognized the validity of electronically signed documents in the country. In 2022, the Ministry of Internal Affairs and the Council of Ministers created clearer rules for the use of digital signatures in government bodies, the Central Bank of Cuba, and local agencies. Currently, Decree-Law 79/2023 and its regulation Decree 106/2024 lay the legal foundations for public key infrastructures in Cuba.

The use of a PKI as an individual component does not guarantee an adequate deployment of the digital signature, in other words, it only supports the issuance of digital certificates that are associated with identity. Key management is of great importance when designing a platform for digital signatures. In this context, asymmetric cryptography is used (Menezes, 1996; Cohen, 2006), so each person or entity will have a key pair: a **public key** that, as its name

indicates, is of public knowledge and a **private key** that must be very well protected. There are both software PKCS#12 (p12, .pfx, jks) and hardware (hardware security modules, HSM) standards that allow, through cryptographic algorithms, to secure the keys (Gitlan, 2025). The public key is used in the **encryption** process and the private key in the information **decryption** process. It is essential to ensure that private keys do not navigate through insecure channels or spaces that facilitate their capture by intruders. In critical infrastructures, it is important to establish robust policies and procedures for key management with high levels of security. Kuzminykh (2021) conducts a detailed study of various tools for key management (KMS, Key Management Service) where criteria such as free or paid software, secure storage, HSM support, use of audit logs, among others, are defined.

Furthermore, tools are needed to carry out the signature transparently and simply. These tools can be diverse: an APK for cellphones, a desktop app or another type of application as long as it has security guarantees. Finally, a digital signature server is essential; it must be equipped with good functionalities and interact with a timestamp service to achieve the validity of a signed document even when the certificate has expired, and in that case, timestamps are essential. In the Cuban environment, different companies make a great effort to develop digital signatures, both for commercial purposes and their business development. Some examples are: XETID, Softel, Datys, Tecnomática and Segurmática that have PKIs certified by MININT (Hernández, 2024). When reviewing each public key infrastructure proposal, the integration of all the aforementioned elements is not fully achieved.

For the aforementioned reasons, this paper proposes an inte-

[3] Decree-Law 79/2023 "On the development, application, and use of cryptographic protection devices and services in the field of cryptography in the Republic of Cuba" (GOC-2024-527-O89)

[4] Decree 106/2024 Regulation on the decree-law 79 "On the development, application, and use of cryptographic protection devices and services in the field of cryptography in the Republic of Cuba", October 26, 2023 (GOC-2024-528-O89)

grated and secure digital signature platform within the context of XETID Comapny. To this end, the platform's general interaction has been designed, and designs and tools have been proposed for each of its components, along with the desired functionalities and security standards to be achieved.

Materials and methods

This proposal is based on the experience learned from developing and deploying Rinho PKI, (developed by the Institute of Cryptography of the University of Havana) in the infrastructure of the XETID Company, based mainly on asymmetric cryptography with elliptic curves (Menezes, 2006) and in harmony with international standards in this field. It should also be mentioned that the main component (the PKI) already has a significant maturity degree and all the resources and infrastructures are available to undertake the proposal made here.

In order to promote the informatization of Cuban society and optimize the services that the XETID Company offers to third parties, there is a need to develop new projects that integrate the digital signature ecosystem. It is worth noting that the proposed functionalities for each component are minimal and can be expanded as needed during development. However, the requirements for key handling set forth here remain strict. This paper focuses on protocols that ensure secure interaction between components without delving into their internal functional details.

The proposal has six essential components:

1. Public Key Infrastructure: a PKI includes the necessary policies, functions, hardware, software and procedures to create, manage, distribute, use, store and revoke digital certificates. Its objective is to facilitate the secure electronic information transfer for a series of network activities, such as: e-commerce, banking and confidential email, among others. The process involves:

- Digital Certificates: electronic credentials that link the identity of the certificate holder to a key pair that can be used to encrypt and sign information.

- Certification Authority (CA): a trusted entity, authorized and certified by MININT, that issues digital certificates. The CA manages all aspects of PKI certificates, including their life cycle.
- Registration Authority (RA): responsible for accepting certificate requests and authenticating the person or organization submitting them.
- Validation Authority (VA): responsible for the certificate validation process based on OCSP (Online Certificate Status Protocol) and CRL (Certificate Revocation List) services.
- Centralized Management Web Application: a web application through which users can request and manage their digital certificates.

2. **XFirma APK:** It is a mobile application that allows users to sign documents and manage their private keys within their device environment. Users may sign documents with their .p12 key file or obtain a certificate from the CA based on a CSR request generated by the XFirma application itself, which is more recommended. It also enables the validation of certificates and documents signed by the APK or by third parties. This type of application has the potential to freely expand its functionalities to ensure document encryption and sharing among users under certain requirements. It can also send encrypted documents via email following a key exchange, among others functionalities.

3. **Signature Server:** This server is responsible for digitally signing documents in various formats, including .pdf, .xml, and code, among others. It primarily uses European standards for digital signatures, such as eIDAS (Regulation of October 9, 2014). It is suitable for use in corporate environments and e-government initiatives. It should offer flexible client interfaces that support integration via REST APIs, web services, the Internet, or through a client-side signing tool integrated with the server. When signing a document, the server manages the signing process and creation of the signed document and integrates the timestamp into it.

4. **TSA Server:** An RFC3161 timestamp server plays a crucial role

in the long-term protection of data records. It provides proof that the data existed at a specific point in time and has not been altered since it was certified and time-stamped—not even a single binary bit. Time-stamping is an essential service of a PKI. It is recommended that the signing certificates used by the TSA server be renewed annually.

5. **Key Management Service (KMS) for digital signatures:** Proper key management and security are essential to the operation of a digital signature server. Therefore, in this proposal, a KMS component is enabled through which users or entities can register and request their digital signature keys. This service may have the following functionalities: complete life cycle management which includes its creation, enablement, disablement, deletion, rotation, modification, and alias assignment; data key management which encompasses creation, encryption, and decryption of data keys; creation and symmetric key management; among other functions.

From the application of the certificate request service using CSR of the CA, the private and public key is generated locally in the KMS and the private key is protected, the CSR request is generated and the corresponding certificate signed by the CA is requested. In this way, the existence of keys for the user or the entity in the KMS is guaranteed. Furthermore, the keys are highly protected. This protection is recommended with the use of an HSM. It is important to point out that in this environment, the necessary cryptographic processes are carried out with the keys stored here (signing, encryption, among others) and requests with the PKCS#11 standard are used, thus guaranteeing that these keys do not come out from the safe environment.

6. **XFirma Corp.:** It is a web application that manages digital signatures for all its registered users or entities, using one or several interfaces of the signature server to manage the signing of documents, XML, or code. In this environment, the utilization of users' .p12 files is not secure, since they would have to travel to the XFirma Corp. server, in addition to trusting an entity not designed for key protection. If the user's intention is to sign with his/her .p12,

then it is safer to use XFirma APK. This web application also allows incorporating functionalities related to the exchange of signed documents among users through prior key exchange or the use of a secure email service.

The use of an identity server for the integration of the components in congruence with their functionalities is essential. This would achieve a single access token for the related components.

For the designs described here, the draw.io application was used, which freely allowed defining the interactions in each of the proposed components.

Results and discussion

Figure 1 shows the design without functional details of the components involved in the aforementioned proposal and their

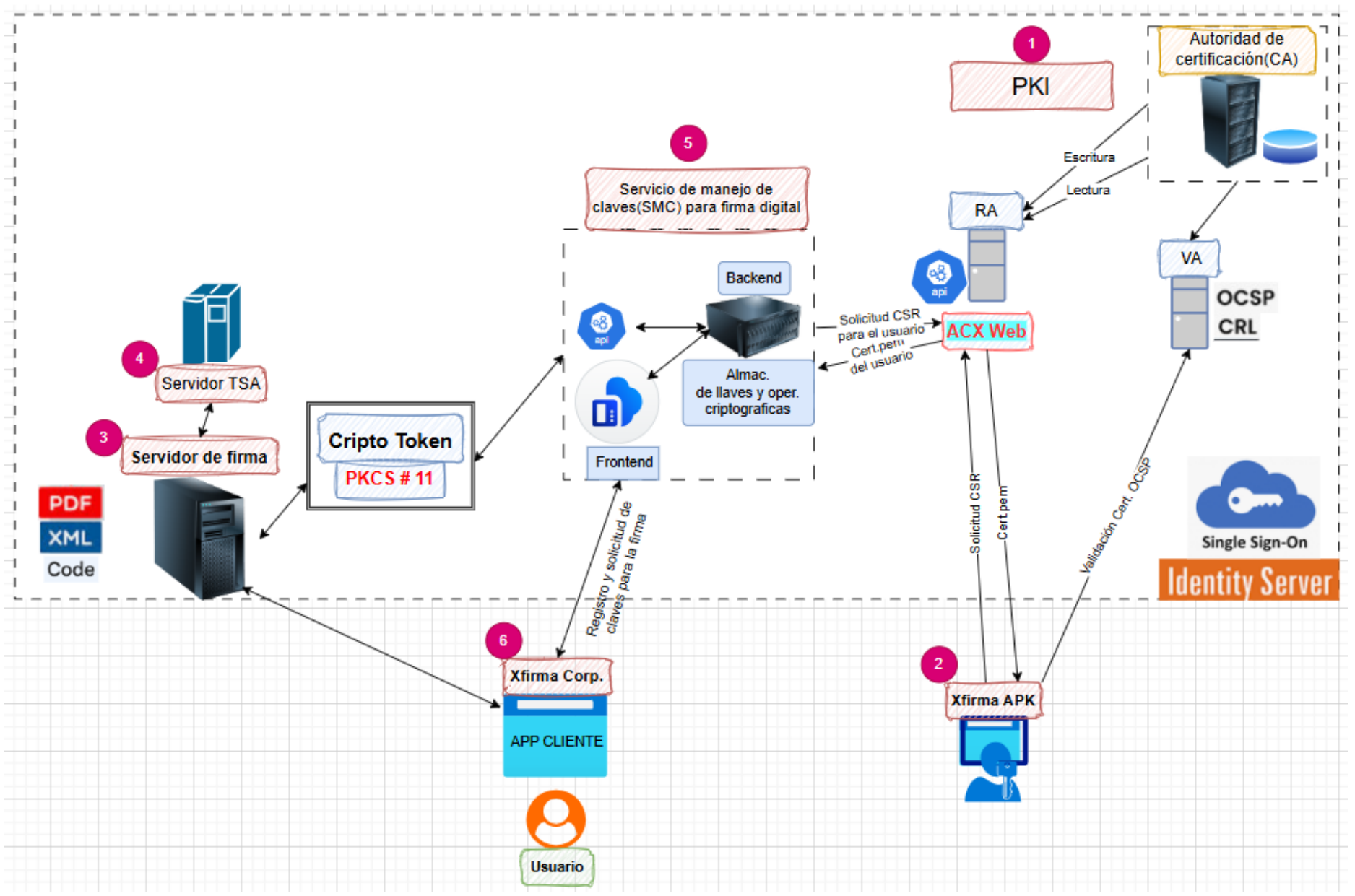


Figure 1. Detailed breakdown of the components and their interaction within the digital signature ecosystem

interrelation.

Breakdown of secure interactions among components and functions related to digital signatures.

1. XFirma APK Interaction or app with CA/RA: by using the Bouncy Castle library with its own random number generator, pub-

lic and private keys are obtained on the user's device. The private key must be kept safe using encrypted files, and then the CSR certificate request is generated and sent to the RA through the defined user interface. Then, it downloads the given certificate in a .pem file signed by the CA. Next, the key container must be built using a .p12 file that will be stored in a secure location by the application. These keys will be associated with the user who previously registered in the application through a registration form adapted to their needs and requirements in their interaction with the PKI. Finally, a document can be signed on the users' side on their devices; the application validates the signature and also the certificate used is validated against the OCSP or using CRL.

2. XFirma Corp. Interaction with the KMS: customers register and create their profile in the KMS, and also requests their corresponding key pair for digital signatures. In this case, the server which has the user's data, is responsible for making the request with a CSR for his/her certificate to the CA. With the certificate and the duly protected private key, the user's .p12 container is created and stored in the KMS. Thus, the user will have his/her registration and the keys for digital signatures on the KMS server. It is important to point out that the KMS can have much more functionalities that are not the objective of this proposal.

3. XFirma Corp. Interaction with the Signature Server: the user requests the signing of a document in his/her profile in XFirma Corp. This web application uses the established user interface and sends to the Signature Server the user ID (IDu), the certificate ID (IDc) and the document to be signed (doc). It is assumed that the user has already created his/her signature profile on the Signature Server with the signature appearance configured.

4. Signature Server Interaction with the KMS: in this case, the Signature Server calculates the hash of the document (h=hash(doc)) sent by XFirma Corp. and sends to the KMS the IDu, IDc and the hash value h of the document using the PKCS#11 standard with the additional order to sign the hash value h. The KMS server, for its part, signs the hash value h of the document with the specified user's keys, finally return-

ing this signed hash to the Signature Server.

5. Signature Server Interaction with the TSA Server: the Signature Server sends the signed hash to the TSA Server for timestamping; the TSA Server adds the timestamp to the signed hash and signs it with its keys. The TSA certificate and the certification chain are also added, thus obtaining the Timestamp Token that is sent to the Signature Server.

6. Signature Server Interaction with XFirma Corp.: now the Signature Server builds the signed document and sends it back to XFirma Corp. Thus, the user obtains the signed document without the need that his/her keys navigate through the infrastructure.

Recommended tools for servers

- **Public Key Infrastructure:** it is widely used in EJBCA business environments from the company KEYFACTOR, with extensive experience in information security developments. This tool has a community and open-source version that allows development on this version. It is programmed in Java and is based on the Bouncy Castle cryptographic library. The Enterprise version of EJBCA has more functionalities but are not free. The use of Rhino PKI is promoted in the Cuban context for a production environment based on EJBCA. It is developed and marketed by the Institute of Cryptography of the University of Havana.
- **XFirma APK:** an Android APK can be developed using a cryptographic library in accordance with the types of documents to be signed. There are several that are free and, if following the PKI development, there is a way to use the Bouncy Castle cryptographic library on Android, thus having a more complete library.
- **XFirma Corp.:** It is a web development that can integrate different technologies and languages, depending on the needs. Bouncy Castle is a library developed for Java and C# languages, therefore, when using JavaScript or Python to carry out the web development, the cryptographic libraries to be used would change. It is also important to consider that libraries are needed for handling digital signatures, validating, and manipulating digital certificates with

the X509 standard. For example, if you decide to use the JavaScript language, in this case a fairly complete library is **node-forge** which is similar to Bouncy Castle. A complete proposal would be:

Backend: Node.js + Express/Fastify, TypeScript (optional, but recommended), Jest for testing.

For cryptography: node-forge / libsodium; bcrypt for passwords; jsonwebtoken for authentication.

For database: PostgreSQL with pgcrypto or MongoDB with field-level encryption.

For frontend: React/Vue/Angular, Web Crypto API for client operations.

- Signature Server: We propose using the community version of KEYFACTOR[5]'s SignServer[6] server, which is open source, and on that version to develop a signature server corresponding to the needs. Figure 2 shows user interaction with the server and the interfaces available for client applications or users in order to enjoy

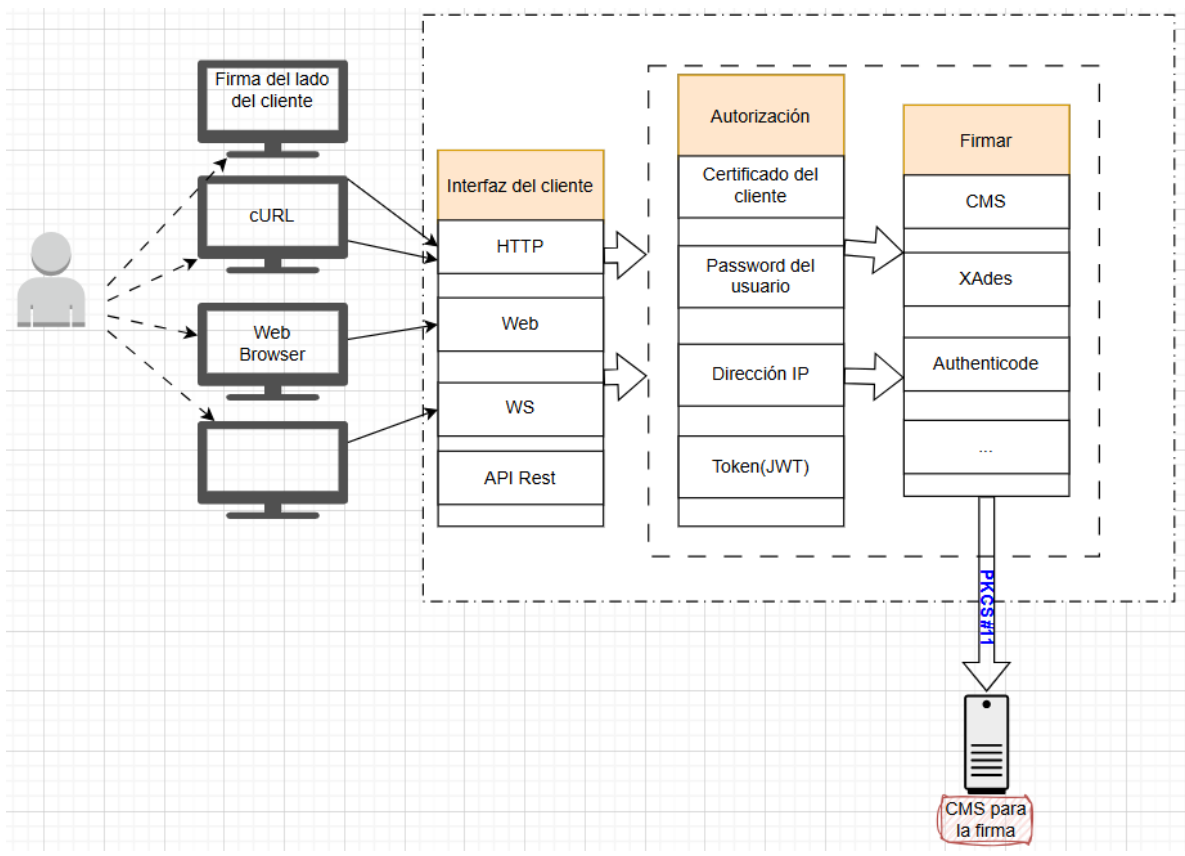


Figure 2. Representation of the Signature Server's functions

the services provided by the Signature Server.

- **TSA Server:** There is an ecosystem of tools that can be used as a TSA service for digital signatures or to program on them. For example: <www.freetsa.org>, <http://timestamp.digicert.com>, OpenTSA, among other options. However, the implementation of

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

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

a TSA server for production requires developments in line with service demand. In the case of businesses, for production it may be advisable to use the TSA server included in the EJBCA platform, which allows working with Docker images, load balancing, and

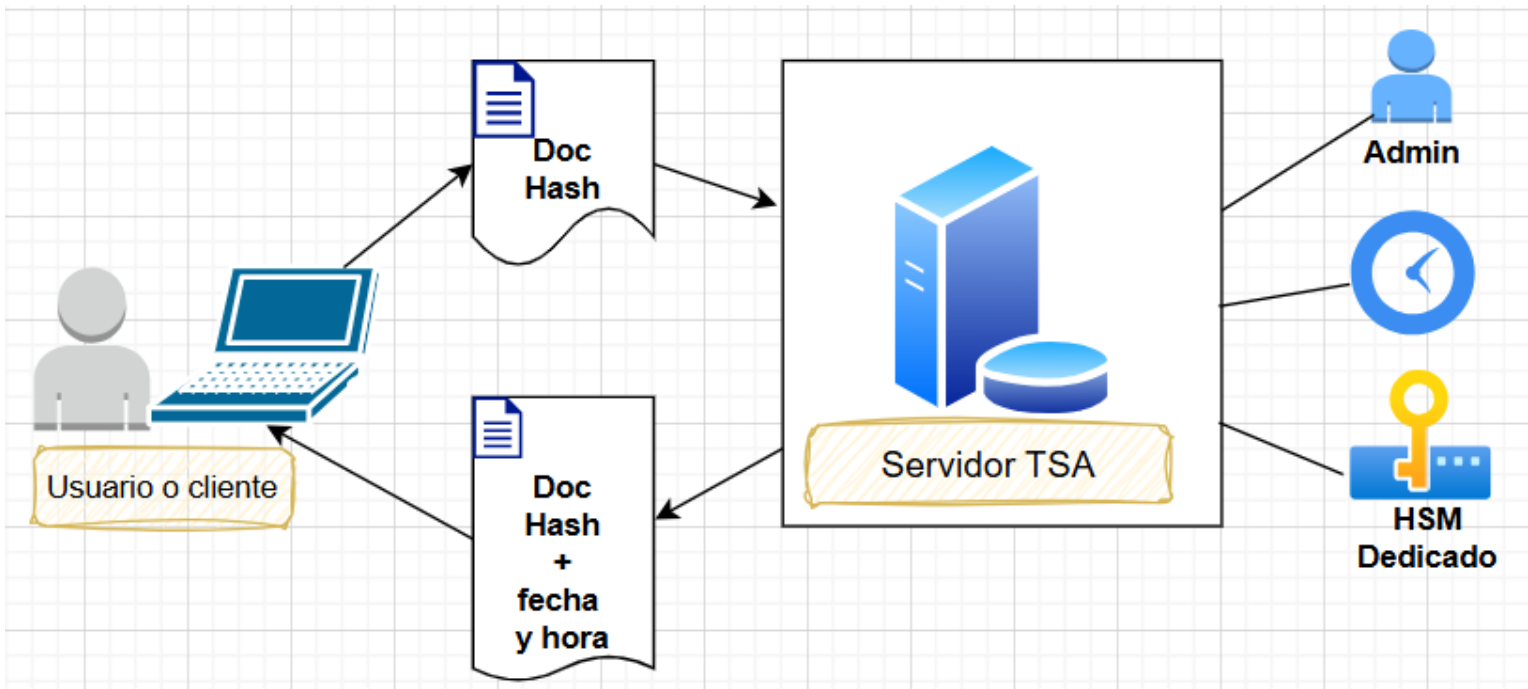


Figure 3. Organizational structure recommended for a TSA Server

adaptation to demand (Figure 3). • Key Management Service (KMS) for digital signatures: there are many cloud-based options available for key management services. Kuzminykh (2021) analyzes a wide range of tools that perform this function with well-defined criteria, enabling users to make an informed decision when choosing one. In the XETID Company environment, it is advisable to carry out in-house development using some of the free services available for this purpose, such as SoftHSMV2, which integrates easily with KEYFACTOR's SignServer and facilitates interaction between the signature server and the SMC Server using the PKCS#11 standard. It is important to point out that this is a workaround for the absence of a professional HSM, but it is not the recommended option. However, if security in the infrastructure and service is maximized, it could be a feasible option.

The development of a platform of this magnitude requires the dedication and commitment of an entire team to achieve secure and user-friendly integration. The modular work of each component is essential for achieving the desired progress.

Conclusions

This research confirms the need for a platform with the potential

to contribute to the informatization of Cuban society through the use of digital signatures. The diversity of existing open-source tools for each one of the described components requires an in-depth review to assess their feasibility for deployment at the production level. In all secure environments, the use of dedicated HSM modules is recommended, but depending on the context, they could be replaced by a similar software solution, which guarantees enhanced security with additional measures, although this is not recommended. It is extremely important that the signature server has different user interface options in order to diversify the types of services offered to users or customers. The company's future projects promote continuity with current developments by integrating the services des-

References

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.

CHRONOLOGY

The Impact of New Technologies on Cybersecurity

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

Received: 11/2025 | Accepted: 11/2025 | Published: 11/2025

Abstract

The development of Information and Communication Technologies (ICTs), together with the advancement of 5G and 6G communication networks, has enabled the evolution of the Fintech market. This market facilitates the information exchange, but has in turn led cybercriminals to steal sensitive information from governments and TELCOS with the aim of causing them to become inoperable and obtaining finance through blackmail. This research examines the growing impact of cybersecurity in the telecommunications sector. The study describes specific threats such as Distributed Denial of Service (DDoS) attacks, banking Trojans, zero-day vulnerabilities, and AI-driven malware. As a result, the impact of new technologies such as AI, the Internet of Things (IoT) and e-commerce has increased vulnerability, as well as the interrelation among regulatory bodies, companies, and governments through regulatory frameworks and the constant training of users as a crucial element to face these threats.

Keywords: artificial intelligence, cyberattacks, TELCOs, cybersecurity, cybercriminals

[*] 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

Introduction

Cybersecurity emerges as an essential discipline, derived from advances in network infrastructure, hardware, software, and the increase in interconnection and digitalization of strategic services such as Banking and Fintech solutions. While these processes generate significant benefits, they also expose organizations to growing risk and make them potential targets of malicious attacks. The objective of such attacks is the information theft, damage to digital assets, or the interruption of operational continuity, with the ultimate goal of generating destabilization, economic losses, and deterioration of customer trust.

The United Nations Office on Drugs and Crime considers cybercrime to be a very complex concept that encompasses a variety of illegal activities. In this sense, it specifies that cybercrime is an act that violates the law and is committed using ICTs to attack networks, systems, data, websites, among others (Ministerio de Relaciones Exteriores de Cuba[3] , 2021). Several authors define cybercrime, considering the above, as a crime committed through Information and Communication Technologies (ICTs), against individuals, companies or governments. Therefore, cybercrime constitutes an unlawful act that is contemplated in criminal law (Benito, 2023).

According to experts from the Information Systems Audit and Control Association (ISACA), cybersecurity is defined as a protective layer for information files. Besides, the term computer security or electronic information security is used to refer to cybersecurity (Infosecurity Mexico, 2025).

It is a fact that the expansion of connectivity comes with an urgent need to protect not only users' personal data, but also the critical infrastructure of companies. In this context, it is indispensable to foster interoperability between telecommunications operators (TELCOs), regulatory agencies, and governments. The objective is to implement joint strategies to prevent, detect, and mitigate cyber threats. Likewise, it is essential to strengthen institutional and citizen capacities

[3] Ministry of Foreign Affairs of Cuba

through continuous training; this measure guarantees comprehensive protection against emerging risks of the digital environment.

Materials and Methods

This research adopts a mixed methodological approach with a qualitative preponderance. The **inductive-deductive** and **analytical-synthetic** methods were used for the information examination. The technique of documentary analysis was essential for the selection and critical evaluation of bibliographic sources, which allowed for a systematic exploration of the subject under study.

Global Cybersecurity Outlook

According to the International Monetary Fund (IMF), cybercrime is estimated to cost globally 23 trillion dollars by the year 2027. In the Latin American region, there has been a 108% increase in cyberattacks on telecommunications operators compared to the previous year, with more than 2,600 weekly incidents per organization. Mexico, Colombia, and Brazil lead the reports in volume and sophistication of threats. The average cost of a data breach was $4.88M in 2024. That same year, ransomware was consolidated as the main threat, both globally and in Latin America and the Caribbean, representing 38% of all registered cyber threats. The main sectors attacked by this threat were wholesale trade and services, health, and construction. The main ransomware with the greatest impact in the Latin American area were RansomHub, LockBit, and Akira. Figure 1 shows a summary of the global impact of this threat.

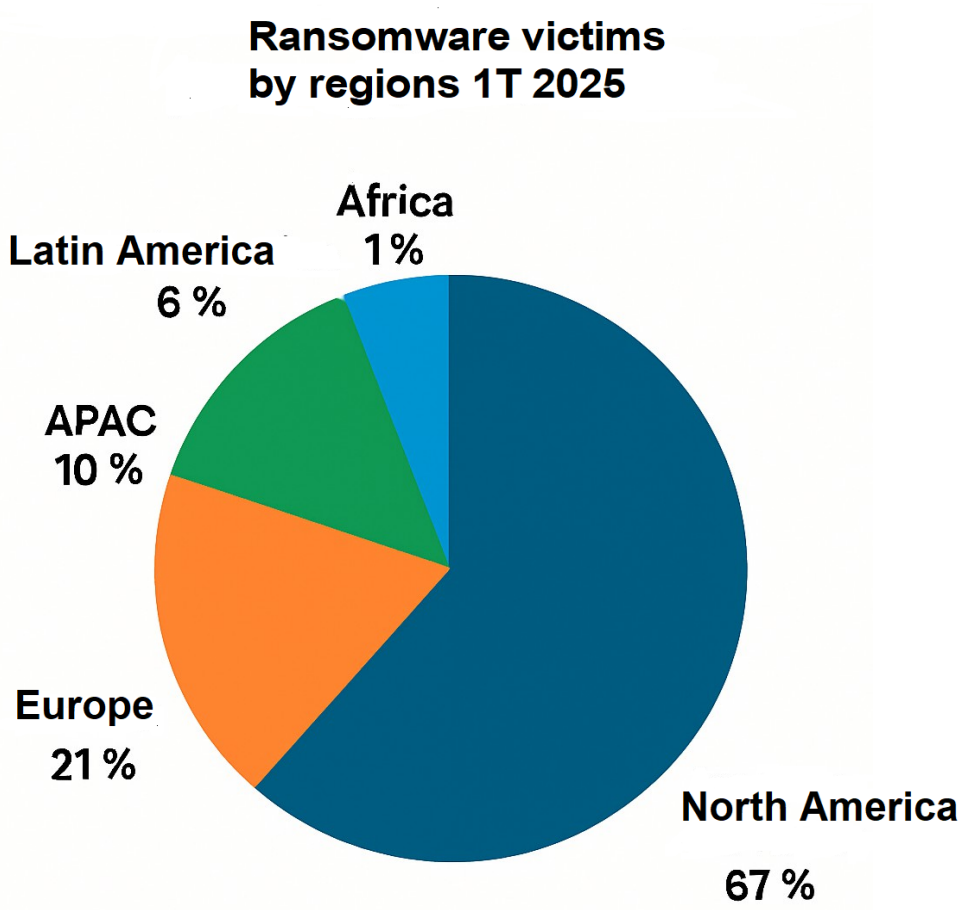


Figure 1. Ransomware victims globally.
Source: Check Point, 2025

The spectrum of threats has expanded considerably and their operations are oriented towards high-impact targets, such as governments, critical infrastructures, and corporations of high strategic value. The main attack vector against these sectors is constituted by Advanced Persistent Threat (APT) groups. These entities are often sponsored by nation-states or sophisticated criminal organizations. They specialize in executing targeted, long-term intrusion campaigns designed to cause maximum strategic impact. Unlike conventional cybercriminals, APT groups operate with a long-term focus. A paradigmatic example of this type of group is Lazarus, an APT sponsored by the North Korean state. Attributed to the Reconnaissance General Bureau (RGB) (the country's main intelligence and special operations body), the group is internationally recognized for the magnitude and global impact of its cyberattack campaigns. It specializes in military espionage, clandestine operations, and offensive cyber activities.

In the year 2024, significant APT attacks have occurred in the world such as:

- MirrorFace (Origin: Japan; scope: Japan and European Union), uses the HiddenFace tool. It employs its attack line through a spear-phishing email where MirrorFace includes a link to a ZIP file. Inside, with an LNK disguised as a Word document that once opened, displays a decoy document and executes version 5.5.5 of the ANEL backdoor.
- Gamaredon (Origin: Russia, scope: Ukraine, NATO), uses the tools: PteroGraphin, PteroPSDoor, PteroSig. Its attack strategy is based on making modifications to one of its PowerShell-written backdoors, known as PteroPSDoor. These improvements included the inclusion of multiple layers of obfuscation and the storage of key components in the Windows© registry, which significantly increased its potential to operate secretly.

"Digital transformation" is the process by which organizations use digital technologies to fundamentally transform the way they operate and deliver value to customers. It is about adopting new tools and methods that can improve efficiency, optimize customer experience, and create new business models. Along with this, challenges can be

mentioned, among which cybersecurity stands out. Companies, to overcome this threat, must rely on a detailed IT incident report template to ensure comprehensive documentation and a rapid response to potential security breaches.

Attacks on Key Sectors

- **Ransomware attack on National Defense Corporation: Date:** March 2025. **Impact:** 4.2 TB of confidential data compromised. The Interlock Ransomware group attacked the National Defense Corporation (NDC) and its subsidiary AMTEC in March, exfiltrating 4.2 terabytes of data that were later leaked on the dark web. **Key Learning:** This incident highlights the importance of ensuring regulatory compliance across the entire supply chain, including internal and external suppliers, in addition to security standards.
- **Cyberattack on WestJet: Date:** June, 2025. **Impact:** Intermittent disruptions to the website and mobile app. In June, Canadian WestJet Airlines revealed a cybersecurity incident that disrupted access to its website and mobile app. **Key Learning:** Organizations must adopt a multi-layered defense that includes regular security training, especially in social engineering tactics. To train employees, especially IT and support teams, in order to recognize identity impersonation attempts can prevent attackers from becoming entrenched.

Cyber Attack Trends

Cyber attacks are based on data hijacking. Hospitals, small and medium-sized companies have been the main victims. Below, see a list of cyber threats that have been evident worldwide against different sectors and are expected to increase in the coming years.

- **Distributed Denial of Service attack or DDoS:** it consists in causing a server to crash by overloading its bandwidth. These actions force the interruption of a website. In the case of the financial system, DDoS attacks are used to flood the online services of banks and trading platforms with a large amount of traffic. In this way, the server collapses and stops working.

- **Banking Trojans:** these are other malicious software that in practice seem harmless, however, they are not and are lurking about the banks. Trojans can be installed on any device by visiting an infected website, downloading an email attachment, or even by downloading an application. Once this virus is installed on the cell phone, it detects when the online services of a bank are used and thus captures personal and banking data.
- **Zero-day vulnerabilities:** attackers look for unknown bugs in programs, IoT devices or cloud services. These gaps allow access to data before developers have a chance to correct them.
- **AI-driven malware:** code generation using artificial intelligence (AI) has significantly democratized and accelerated the development of malware. Attackers can generate, refine, and obfuscate malicious software variants with unprecedented efficiency, facilitating evasion of signature-based and heuristic detection mechanisms.
- **Exploitation of insecure APIs:** poorly configured or unprotected application programming interfaces (APIs) provide a window to extract data, as many companies rely on third-party services with weak security controls.
- **IoT devices as a gateway:** sensors, cameras, peripherals, and other devices connected to the corporate network can be used as attack vectors to access internal systems and extract data.

Blockchain, Cloud, Telework

Blockchain

Technologies such as blockchain appear as solutions to improve data security and trust. Cyberattacks targeting cloud, hardware, and machine learning systems can create complex breaches, making it difficult for companies to protect their data. Blockchain technology, with its transparent and immutable books, offers innovative solutions for cybersecurity.

Blockchain audits can address various attacks and vulnerabilities in smart contracts. For threat response management, blockchain enables automation and decentralization, allowing for rapid identification, containment, and mitigation of threats without compromising

data security. Smart contracts can respond to cyberattacks by isolating threats automatically and quickly. From a cybersecurity perspective, blockchain technology introduces transformative advances: it strengthens operational resilience, reinforces access control models, and guarantees data integrity. Its decentralized architecture mitigates the impact of DDoS attacks by eliminating single points of failure, making service disruption significantly more difficult.

At the same time, Zero Trust principles are enhanced through native identity verification mechanisms and granular access control, reducing the risks of unauthorized access and insider threats. Additionally, decentralized file storage (based on distributing encrypted fragments across multiple nodes) minimizes exposure to data breaches, while the inherent immutability and replication of the network ensure continuous data availability and integrity, even during security incidents.

Cloud

Cloud security is a cybersecurity discipline dedicated to securing cloud computing systems. It includes keeping data private and secure across online infrastructure, applications, and platforms. At its core, cloud security is made up of the following categories:

- Data security.
- Identity and access management (IAM).
- Governance (policies for preventing, detecting, and mitigating threats).
- Data retention (DR) and business continuity (BC) planning.
- Legal compliance.

Cloud security is all the technology, protocols, and best practices that protect computing environments, the applications that run on them, and the data stored in them. Service security starts by understanding what exactly is being ensured, as well as the aspects of the system that need to be managed. The total scope of cloud security is designed to protect the following:

- **Physical networks:** routers, electrical power, cabling, climate controls.
- **Data storage:** hard drives.

- **Data servers:** computer hardware and software of the core network.
- **Computer virtualization platforms:** virtual machine software, host machines, and guest machines.
- **Operating systems (OS):** software that supports all computer functions.
- **Middleware:** application programming interface (API) management.
- **Execution environments:** running and maintaining a running program.
- **Data:** all information stored, modified, and accessed.
- **Applications:** traditional software services (email, tax software, productivity packages).
- **End-user hardware:** computers, mobile devices, IoT devices.

Telework

Telework was consolidated as a fundamental pillar for business continuity during the pandemic, allowing countless organizations to operate. However, this modality has critically expanded the attack surface, transforming remote employee access into a threat vector frequently exploited by malicious actors. To face this risk landscape, the immediate adoption of a set of best practices about security designed to strengthen defenses and minimize exposure to cyberattacks is imperative; these are described below:

- **Defining a security policy for telework:** it is strictly necessary to create a policy that defines the protocols, order, and crisis plan for each of the scenarios in which the user may be involved.
- **Using the Zero Trust model:** based on security guidelines, solutions, and policies to protect the corporate work environment.
- **Managing device administration:** working with a cloud-based management environment and apply appropriate policies for its access.
- **Using multi-factor authentication (MFA):** having a solid identity base that allows users to securely access resources and applications wherever they are.

- **Working in a secure environment:** accessing resources such as servers or connect to company resources via VPN.
- **Implementing integrated cloud security solutions:** connecting all identity, application, and data solutions into a single ecosystem.
- **Conducting periodic safety testing:** carrying out regular safety audits as a requisite.
- **Performing automated backups:** performing backups as a measure of prevention and information security to ensure continuity against any incident that may affect the systems.

Cybersecurity in 5G and 6G Networks

The current 5G network and its evolution towards 6G allow for higher connection speeds, lower latency, and new possibilities for massive interconnection between IoT devices, as well as the development of interactive holograms in real-time and sensory Internet, all powered by artificial intelligence. However, these evident technological advantages come with parallel security risks. Therefore, all sectors of society must prepare to achieve the necessary resilience against cyber threats and attacks on their computer systems.

Some of the risks associated with its deployment include:

1. **Larger attack surface:** 6G technology promises the connectivity of more devices, which translates into an increase in potential vulnerabilities to cyberattacks.
2. **Privacy compromise:** this technology includes more precise real-time location and tracking adjustments, and this compromises user security.
3. **More advanced attacks:** 6G networks need more AI to strengthen, but, as in the case of these tools currently in use, it can help cybercriminals carry out attacks that are harder for users to detect.
4. **Vulnerability of IoT devices:** by including a large number of new IoT devices in their scope, they can be more easily subject to cyberattacks if they are not adequately protected.

General Strategies Against Cyberattacks

TELCOs and companies must implement key cybersecurity measures to protect their operations and customer trust. Important recommendations to keep in mind are listed below:

- **Implement advanced detection and response systems:** detecting threats early is key to preventing further damage. Advanced detection and response systems identify unusual patterns in network traffic and respond quickly to potential attacks. These tools, based on AI and machine learning, are useful for detecting unknown threats.
- **Adopt multi-factor authentication:** passwords alone are not enough to protect access to sensitive systems. Multi-factor authentication adds an extra layer of security, ensuring that only authorized personnel access critical information, something essential in the context of increasingly common remote access.
- **Instruct employees in security practices:** social engineering attacks, such as phishing, take advantage of the team's lack of knowledge to access systems and data. Instruct staff in detection of suspicious emails and passwords protection is key to preventing security breaches.
- **Regularly update and patch all systems and devices:** one of the most common mistakes is not keeping systems updated. Software updates include patches that fix known vulnerabilities. It is key for telecommunications companies to establish a regular maintenance schedule to reduce the risk of attacks.
- **Implement strict access and permission policies:** limiting access to systems and data based on role is a good cybersecurity practice. Companies must ensure that only authorized personnel access confidential information and adjust permissions according to each employee's functions.
- **Conduct periodic security audits and assessments:** cybersecurity evolves and threats change. Audits and assessments help identify vulnerabilities and adjust security measures, in addition to verifying the effectiveness of implemented policies and tools. Cybersecurity is essential in telecommunications. In an environment where

connectivity and data exchange are fundamental, protecting information and infrastructure maintains customer trust and ensures regulatory compliance. Robust measures and attention to new threats strengthen any company in the market.

Impact of AI on Cybersecurity

AI has established itself as a powerful tool that defenders use to anticipate, detect, remediate, and mitigate security incidents more effectively. However, cybercriminals also use it to automate vulnerability search, generate more sophisticated phishing attacks, and create malware.

The convergence of Fintech technologies; IoT with faster networks, such as 5G, will allow a much higher volume of traffic than currently. Therefore, it is necessary to project and use these new capabilities to defend systems against cyberattacks. Below, see the key elements to consider for the use of AI in this context.

Smarter and Harder-To-Detect Attacks

- Hyper-realistic phishing: with AI, an attacker can analyze social profiles, emails, images, and the language used by a victim to build fraudulent messages almost indistinguishable from legitimate ones.
- Automation of vulnerability reconnaissance: previously, a cybercriminal manually examined systems to find vulnerabilities. Now, with AI, they can analyze thousands of components on a large scale and in record time to identify vulnerabilities and weaknesses.

Recommendations

- Security tools with integrated AI: adopting solutions based on machine learning and behavior analysis to detect threats in real-time.
- Indispensable human supervision: analysts with solid knowledge and experience in cybersecurity interpret the results of AI, identify false positives and negatives, and make informed decisions. AI is not infallible and the human factor remains essential.
- Work with specialized technology partners: Many companies opt for managed security services (MSSPs) that use AI for monitoring

and incident response, taking advantage of the capabilities and expertise of cybersecurity specialists without having to invest in technology, training, and hiring an in-house team.

Regulatory Framework for Cybersecurity

1) **The Budapest Convention on Cybercrime:** this convention is formally known as the Convention on Cybercrime of Council of Europe and was opened for signature on November 23rd, 2001, as the first international treaty addressing crimes committed through the Internet and other computer networks.

2) **The General Data Protection Regulation (GDPR):** this regulation establishes strict standards for the processing and storage of personal data in the European Union (EU).

3) **ITU-T X.1157 Recommendation, 2015:** it describes the necessary capabilities for the fraud detection and response service for ICT-based services.

4) **Health Insurance Portability and Accountability Act (HIPAA):** this U.S. law came into effect in 1996 and aims to protect the privacy and security of information within the healthcare sector.

5) **Payment Card Industry Data Security Standard (PCI DSS):** it is a set of security standards and requirements designed to protect payment card information and ensure the security of credit and debit card transactions.

6) **ISO 27001:** it is an international rule that focuses on information security, providing a framework for establishing, implementing, maintaining, and improving an information security management system (ISMS).

7) **NIST Cybersecurity Framework:** it is a set of standards, guidelines, and best practices designed to help organizations manage and improve their cybersecurity. It is based on three fundamental principles: risk management, flexibility, and scalability, structured into five functions:

- Identifying vulnerabilities and security gaps in a computer system.
- Protecting computer systems.

- Detecting malicious activities in a computer system.
- Knowing how to respond to a computer incident.
- Recovering information after a computer incident with full capacity and resources.

8) Cuban Provision on Cybersecurity: Resolution 105, the Regulation on the National Action Model for Cybersecurity Incident Response is the implementation of the provisions of Decree 360/2019. It is preventive and reaches the whole society, by implementing a working system between the specialized entities in ICT security for the fulfillment of their functions in the secure exchange of information related to vulnerabilities and cybersecurity incidents.

Investment in Cybersecurity

Different factors that impact and constitute risks, trends, and projections in the context of cybersecurity have been addressed. Technological advances require investment in cybersecurity in telecommunications companies, which is not just about protecting assets, but also about ensuring resilience and sustainability, by focusing on key areas such as risk assessment, supplier management, business continuity planning, incident response, and continuous testing. Investment should focus on strengthening cloud services, blockchain, quantum computing, and the use of AI tools that allow protecting telecommunications infrastructures from threats or attacks.

Cases of Cyberattacks on TELCOs During This Year

Telefónica

In January 2025, Telefónica suffered a significant cyberattack that resulted in the leak of approximately 2.3 gigabytes of internal data. The attack originated from a sophisticated phishing campaign targeting Telefónica's employees. Cybercriminals used social engineering techniques to deceive employees and obtain their access credentials. Once inside the system, they accessed the company's internal ticketing system, known as Jira, and extracted a significant amount of data.

Consequences of the Cyberattack

1. **Internal and corporate customer data leakage:** More than 236,000 business customer records and around 469,000 internal ticket records were exposed, raising privacy and security concerns.
2. **Reputational damage:** the exposure of sensitive information affected Telefónica's image and diminished the trust of its clients and strategic partners.
3. **Potential legal sanctions:** the data leakage can lead to financial sanctions due to non-compliance with regulations such as the General Data Protection Regulation (GDPR).
4. **Increase in fraud attempts:** cybercriminals can use the leaked information to carry out targeted attacks on affected customers.

Response

Once the attack was detected, Telefónica activated its incident response protocol and worked with its cybersecurity team to mitigate the impact and prevent further leaks.

1) Compromised accesses were identified and blocked. This action made it possible to isolate the accounts used by the attackers and reinforce access controls.

2) The mandatory reset of credentials for all employees was established. This measure was accompanied by the implementation of stricter multi-factor authentication mechanisms.

At the same time, the Company carried out a forensic analysis of the attack with the aim of determining which vulnerabilities were exploited and how to improve its security infrastructure. As a result of this investigation, Telefónica decided to strengthen its internal defenses, implement real-time threat detection solutions, and reinforce network segmentation to hinder the mobility of potential attackers within its systems.

Lessons Learned and Prevention Strategies

The incident highlights the importance of adopting a proactive approach to cybersecurity. Some key strategies to prevent similar attacks include:

- **Multi-factor authentication (MFA):** prevent credential theft from being enough to compromise critical systems.

- **Cybersecurity awareness:** train employees to detect phishing attempts and apply good security practices.
- **Advanced monitoring:** use artificial intelligence and behavior analysis to detect unusual accesses.
- **Access management with minimum privileges:** limit permissions to what is strictly necessary for each user.
- **Periodic penetration testing:** evaluate system security through real attack simulations.

Claro Argentina

On July 4, 2025, attackers stole the user database of Claro, one of the three main telecommunications companies in Argentina. The stolen information is currently for sale on the dark web.

How the Attack was Carried Out

The attack targeted the Claro app and website, where users manage their services and make purchases. This means that cybercriminals did not access the cell phone lines themselves, but they did access accounts associated with those services, from which it is possible to make equipment purchases, request services, and manage personal information.

The leak includes emails, full names, addresses, national ID numbers (ID), and zip codes, sensitive data that is already circulating on the deep web and was also uploaded to sites like BiteBlob.com, a platform that allows anonymous file sharing. In these forums, data is sold to the highest bidder or even published for free, putting the privacy of thousands of people at risk.

Actions Taken by Claro

Internal analysis protocols were activated after detecting that the company was being mentioned in forums where stolen databases are often offered.

Recommendations

Experts generally recommend a series of measures for users, including:

- Change passwords regularly and do not reuse them across services.

- Activate two-step verification on all accounts.
- Avoid entering personal data on unofficial or suspicious sites.
- Be alert to possible phishing emails and do not download unknown files.
- Check with companies if your accounts have been compromised and follow official instructions.

Bouygues Telecom

France's third-largest telecommunications carrier, with a base of around 27 million users, confirmed on August 4 that it had been the victim of a cyberattack. The incident resulted in a data breach that compromised the personal information of approximately 6.4 million of its customers, equivalent to almost a quarter of its total user base in the country.

Consequences

The attackers were able to obtain contact information, contract details, and their International Bank Account Number (IBAN); affected individuals include private customers and companies. However, it appears that the most confidential data, such as passwords and credit card numbers, were not affected by this breach.

Actions

The carrier filed a complaint with the judicial authorities. In it, it stated that the perpetrator could face up to five years in prison and a fine of 150,000 euros. The TELCO has not explained the type of attack, nor the vulnerability attacked in its network.

Orange France

The French telecommunications multinational Orange has confirmed that it was the victim of a cyberattack on July 25, 2025.

Actions

The attack forced its security team, in collaboration with Orange Cyberdefense, to take immediate containment measures. The incident, which affected one of its information systems, had consequences on various services for businesses and consumers, especially in France.

Measures

The company filed an official complaint on July 28, 2025, while authorities are already involved in the investigation. It does not

mention any evidence of exfiltration of internal or customer data. As soon as the attack was identified, Orange's technical teams mobilized to isolate services that may have been compromised. This preventive action had the collateral effect of disrupting some management platforms for business customers and certain services for individual users.

Results and Discussion

The results obtained show a growing vulnerability of the telecommunications sector to increasingly sophisticated cyber threats. In Latin America, there has been a 108% increase in weekly incidents, with Mexico, Colombia, and Brazil being the most affected countries. This data reveals an alarming trend that requires a deep review of digital security policies in critical infrastructures.

Among the most frequent threats, ransomware stands out as the predominant attack vector, accounting for 38% of global incidents. Variants such as LockBit and Akira have demonstrated a significant penetration capacity, affecting both large corporations and regional providers. Furthermore, an increase in the activity of APT groups, such as Lazarus and MirrorFace, which operate with strategic objectives and state backing, is observed, adding a geopolitical dimension to the problem.

The research shows the role of emerging technologies such as AI in the evolution of attacks. AI is being used to generate more evasive malware and automate hyper-realistic phishing campaigns. In turn, the expansion of 5G and 6G networks, along with the exponential growth of IoT devices, has expanded the attack surface, generating new vulnerabilities in APIs and authentication systems.

The analysis of real cases of attacks on Telefónica, Claro Argentina, and Bouygues Telecom illustrates how security breaches can compromise the privacy of millions of users and damage corporate reputation. The responses implemented by these companies, such as multi-factor authentication (MFA), network segmentation, and forensic analysis, constitute valuable examples of good practices that can be replicated in other contexts.

Regarding the regulatory framework, the importance of instruments such as the Budapest Convention, the General Data Protection Regulation (GDPR), the ISO 27001 standard, and Resolution 105 is highlighted in Cuba. International cooperation and interoperability among governments, telecommunications operators, and regulatory bodies are emerging as key elements for an effective and coordinated defense. Furthermore, the dual role of AI in this scenario is evident. On one hand, it allows detecting threats in real-time and optimizing incident response. On the other, it poses ethical and operational risks by facilitating automated attacks. In this context, human supervision remains essential to interpret data and make strategic decisions.

Together, the results confirm that cybersecurity is no longer a merely technical issue but has become an essential component of business and national strategy. The research provides a comprehensive view of the current landscape and suggests that future research should focus on the economic impact of cyberattacks and the evaluation of AI-based solutions from an ethical and sustainability perspective.

Conclusions

Cybersecurity is consolidated as a strategic axis in the telecommunications sector, driven by the rapid adoption of emerging technologies such as Fintech, IoT, artificial intelligence (AI), quantum computing, and 5G networks. This increasingly interconnected digital ecosystem has significantly expanded the attack surface, leading to more sophisticated and persistent threats.

In this context, AI plays a dual role: on one hand, it enhances defensive capabilities through predictive analysis and response automation; on the other, it enables new, more adaptive and harder-to-detect attack vectors. Faced with this panorama, it becomes imperative to strengthen the technical capabilities of carriers, foster interoperability among the different actors of the digital ecosystem, and promote a resilient organizational culture, capable of anticipating and mitigating risks. Interinstitutional collaboration, together with continuous professional development, is positioned as an essential component to face a dynamic, distributed, and constantly evolving threat environment.

References

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/