Romanian Association for
Information Security Assurance

PROCEEDINGS
OF
THE INTERNATIONAL CONFERENCE ON CYBERSECURITY AND CYBERCRIME

Volume XII
eISSN 2393-0837

CyberCon Romania
2025

THE INTERNATIONAL CONFERENCE ON CYBERSECURITY AND CYBERCRIME

PROCEEDINGS
OF
THE INTERNATIONAL CONFERENCE ON CYBERSECURITY AND CYBERCRIME

Volume XII

A scientific conference organized by the
Romanian Association for Information Security Assurance

CyberCon Romania
2025

THE INTERNATIONAL CONFERENCE ON CYBERSECURITY AND CYBERCRIME

The International Conference on Cybersecurity and Cybercrime (IC3) is an annual scientific conference, with the purpose to encourage the exchange of ideas about the evolution of cyberspace, information security challenges, and new facets of the phenomenon of cybercrime. The event provides the appropriate framework for experts to present their research in this field.

The Proceedings of the International Conference on Cybersecurity and Cybercrime includes scientific papers reviewed by the *Editorial Board* that consists of experts from the academic field, their work taking place under the guidance of the *Advisory Board*, composed of internationally recognized personalities from the academic field.

Proceedings of the International Conference on Cybersecurity and Cybercrime
Online ISSN: 2393-0837
Print ISSN: 2393-0772
DOI: 10.19107/CYBERCON
URL: proceedings.cybercon.ro

The International Conference on Cybersecurity and Cybercrime is part of the **CyberCon Romania** event, organized by the Romanian Association for Information Security Assurance.

CyberCon Romania brings together experts from public institutions, private companies, and universities, for raising the level of awareness and embodies the cybersecurity culture.

Website: www.cybercon.ro

The Romanian Association for Information Security Assurance (RAISA) is a professional, non-governmental, non-partisan political, nonprofit, and public benefit association.

Founded in 2012, the association started as an initiative with the aim of promoting and supporting information security activities in compliance with applicable laws and creating a community for the exchange of knowledge between the experts from the public, private, and academic environment. Its vision is to encourage the cybersecurity research and education, and to contribute to the creation and dissemination of knowledge and technology in this domain.

Website: www.raisa.org

CONFERENCE COMMITTEES

CONFERENCE CHAIRMAN

Professor Emeritus **Ioan C. BACIVAROV**, PhD
National University of Science and Technology POLITEHNICA Bucharest, Romania
Faculty of Electronics, Telecommunications and Information Technology

INTERNATIONAL SCIENTIFIC COMMITTEE

Professor Emeritus **Alessandro BIROLINI**, PhD
ETH Zurich, Switzerland

Professor **Angelica BACIVAROV**, PhD
POLITEHNICA Bucharest, Romania

Irina BAKHAYA, PhD
"Al. I. Cuza" Police Academy, Romania

Natalia BELL, D.Sc.
Marymount University, United State of America

Professor **Răzvan BOLOGA**, PhD
University of Economic Studies, Romania

Assoc. Prof. **Darius BUFNEA**, PhD
Babeș-Bolyai University, Romania

Assoc. Prof. **Ciprian CONSTANTIN**, PhD
"Al. I. Cuza" Police Academy, Romania

Viorel GAFTEA, PhD
Romanian Academy, Romania

Alexandru GEORGESCU, PhD
National Institute for R&D in Informatics, Romania

Assist. Prof. **Cătălin GOLOP**, PhD
"Al. I. Cuza" Police Academy, Romania

Assoc. Prof. **Niculae IANCU**, PhD
Constanța Maritime University, Romania

Assoc. Prof. **Andrei IGNAT**, PhD
"Al. I. Cuza" Police Academy, Romania

Angela IONIȚĂ, PhD
Research Institute for Artificial Intelligence, Romania

Assist. Prof. **Andrei LUCHICI**, PhD
Romanian-American University, Romania

Assoc. Prof. **Dumitru-Iulian NĂSTAC**, PhD
POLITEHNICA Bucharest, Romania

Pierluigi PERRONE, PhD
LUISS University, Rome, Italy

Professor **Florin POPESCU**, PhD
"Carol I" National Defense University, Romania

Assoc. Prof. **Eduard-Cristian POPOVICI**, PhD
POLITEHNICA Bucharest, Romania

Assoc. Prof. **Ciprian PUNGILĂ**, PhD
West University of Timișoara, Romania

Assoc. Prof. **Gabriel RAICU**, PhD
Constanța Maritime University, Romania

Professor **Răzvan RUGHINIȘ**, PhD
POLITEHNICA Bucharest, Romania

Professor **Anurag SHARMA**, PhD
GNA University, India

Assoc. Prof. **Emil SIMION**, PhD
POLITEHNICA Bucharest, Romania

Professor **Pradeep Kumar SINGH**, PhD
University of Information Technology, India

Assoc. Prof. **Adrian STERCĂ**, PhD
Babeș-Bolyai University, Romania

Assoc. Prof. **Mihai SUCIU**, PhD
Babeș-Bolyai University, Romania

ACKNOWLEDGEMENT:

This volume was created with the support of professors from:
- The Department of Electronic Technology and Reliability, Faculty of Electronics, Telecommunications and Information Technology, National University of Science and Technology POLITEHNICA Bucharest.
- The Department of Police, Faculty of Police, “Alexandru Ioan Cuza” Police Academy.
- The Department of Information Systems and Cyber Operations, Faculty of Security and Defense, “Carol I” National Defense University.

TABLE OF CONTENTS

CyberCon Romania 2025 Welcome Message

Professor Emeritus Ioan C. BACIVAROV, PhD
President of the Romanian Association for Information Security Assurance (RAISA)

Ladies and gentlemen, distinguished guests, colleagues in the field of cybersecurity,

First of all, I would like to extend to you, as President of the *Romanian Association for Information Security Assurance (RAISA),* a warm welcome to the *CyberCon Romania 2025* conference.

As you probably know, I am Professor Emeritus in the field of dependability (reliability, maintainability, security, survivability a.o.) at the Faculty of Electronics, Telecommunications and Information Technology of the National University of Science and Technology POLITEHNICA Bucharest, the most important technical university in Romania. In this capacity, I have introduced and developed several innovative educational and scientific programs in these fields over the last four decades.

I would like to begin by expressing our gratitude to the Embassy of the United States of America in Romania for the constant support provided to the *CyberCon Romania* conference organization over the last years. It is important to mention that this partnership - started at the initiative of the US Embassy- has obtained the *Cyber Outstanding Security Performance Award (Cyber OSPA) 2022* in the *Outstanding Cyber Security Partnership* category.

Special thanks are due to the educational partners of *CyberCon Romania 2025,* including the *National Institute of Research and Development in Informatics - ICI Bucharest,* the *Euro-Atlantic Resilience Center*, the *Romanian Banking Institute,* the *European Institute of Romania,* the *Romanian Institute of Financial Studies*, to the universities involved including the *National University of Science and Technology POLITEHNICA Bucharest*, *"Alexandru Ioan Cuza" Police Academy, "Carol I" National Defense University a.o.,* as well as to all the other organizations and companies involved in this important area participating in the conference, whose names are mentioned on the conference website.

The fact that the work of this conference is taking place at the headquarters of the *European Commission Representation in Romania* wants to emphasize the importance given by the *European Union* to cybersecurity issues. Therefore, we would like to thank, once again, the European Commission Representation in Romania for the support provided in the organization of this conference. Our thanks also go to the media partner of this event, the *National Press Agency AGERPRES*, for the way it reflected RAISA's activities aimed at improving cybersecurity culture in Romania.

Special thanks are due to all the important experts in the field who will present their views during the conference and will thus contribute to shaping a realistic picture of the challenges, risks and solutions in the field of cybersecurity, in the very difficult international period we are going through.

In today's hyperconnected world, cybersecurity has evolved from a technical concern to a foundational pillar of global stability. The modern threat landscape is more complex than ever, shaped by geopolitical tensions, AI-driven attacks, and an expanding digital ecosystem that blurs the line between personal, corporate, and national security.

It is important to note that, in the current complicated international circumstances, the *North Atlantic Treaty Organization (NATO)* treats *cybersecurity* as a core element of its collective defense and deterrence strategy. As we navigate 2025, organizations face a dual challenge: defending against increasingly sophisticated adversaries while maintaining trust, privacy, and operational resilience.

The rise of generative AI, deepfake technology, and quantum computing threats is redefining what it means to be "secure." Meanwhile, human factors - awareness, behavior, and ethics - remain the most critical layer of defense. The path forward demands collaboration, transparency, and continuous innovation. Cybersecurity is no longer just an IT issue - it's a shared responsibility across sectors, governments, and individuals. Together, we must shift from reactive defense to proactive resilience, building a safer digital future for all.

The *Romanian Association for Information Security Assurance (RAISA)* - the main organizer of this conference - is a professional, nonprofit and public benefit association, founded 13 years ago, as an initiative dedicated to disseminating the concept of cybersecurity and fighting against cybercrime. The vision of *RAISA* is to promote research and education in information security field, to contribute to the creation and dissemination of knowledge and technology in this domain and to create a strong "*cybersecurity culture*" at national level.

We are aware that the achievement of a strong "*cybersecurity culture*" can be obtained only through the joint effort of several entities from the public, private and academic sectors. Romania has a very good position in Europe in the field of cybersecurity, and the competence of Romanian specialists in the field is recognized and appreciated. The *cooperation* of all the factors involved is the key word, if the development of a strong and efficient cybersecurity system is desired at national level.

And conferences like *CyberCon Romania* can facilitate the exchange of ideas and the establishment of partnerships in this important field. As technology continues to evolve, so do the threats we face - making collaboration and knowledge-sharing more vital than ever. This conference is not only a forum for ideas but also a community dedicated to building a safer, more resilient digital world.

CyberCon Romania 2025 takes place within a managerial part, structured in several panels, and within a scientific part. During the conference, challenges and future changes in the field of cyber-security will be analyzed.

Finally, we would like to thank all the authors who submitted their work, all experts participating in the panels, and all those who contributed to the Conference with their efforts and support. We hope that the *CyberCon Romania 2025 Conference* will be a both stimulating and enjoyable forum in which to share current results and trends in cybersecurity, to develop new partnerships and we wish it every success.

Let's make together this conference an inspiring and impactful experience for everyone. More information about CyberCon Romania 2025 can be found on the webpage *https://www.cybercon.ro.*

The Fight Against Terrorism in the Digital Era: Policing Perspectives, Legislative References, and Cybercrime Dimensions

Ștefan-Gabriel DASCĂLU, Marius-Andrei OROȘANU
Police Academy "A.I. Cuza" Bucharest, Romania
stefan.dascalu@academiadepolitie.ro, andrei.orosanu@academiadepolitie.ro

Abstract

The digital era has transformed both the nature of terrorism and the mechanisms designed to combat it. Contemporary terrorist organizations increasingly exploit cyberspace for recruitment, propaganda, financing, and operational coordination, thus blurring the boundaries between physical and virtual threats. This paper examines the evolving role of law enforcement agencies in addressing these challenges, emphasizing the need for advanced technological tools, interagency cooperation, and continuous adaptation of policing strategies. Furthermore, it analyzes the legislative frameworks that underpin counterterrorism policies within the digital domain, highlighting existing gaps and the growing intersection with cybercrime. By integrating legal, operational, and technological perspectives, the study aims to provide a comprehensive understanding of how digitalization reshapes both terrorism and the institutional responses designed to counter it.

Index terms: counterterrorism, cybercrime, digital foreign policy, European Union, law enforcement cooperation

1. Introduction

Terrorism represents one of the most serious threats to international security, public order, and the democratic values of the contemporary world. From large-scale attacks such as those of September 11, 2001, to targeted assaults carried out in European capitals - Paris, Brussels, London, Madrid - the terrorist phenomenon has continuously adapted, diversifying its methods, targets, and operational means. Today, terrorist organizations exploit technology, online radicalization, failed-state environments, organized crime networks, and global mobility, which confer upon the phenomenon a fluid, unpredictable nature that is increasingly difficult to counter through traditional means [1].

Transformations in the security environment have compelled states to reorient toward a modern policing model grounded in prevention, intelligence, integrated action, and international cooperation. Whereas traditional terrorist structures once operated hierarchically with visible leaders, command centers, and clear chains of command current trends reveal a shift toward fragmented entities: autonomous cells, "lone wolves," multi-nodal networks, and hybrid groups with ideological, criminal, and paramilitary dimensions. This evolution obliges institutions responsible for counterterrorism to develop proactive capabilities, adopt flexible investigative methods, and rapidly exploit operational intelligence [2].

Across Europe, the rise in high-impact attacks such as the Bataclan (2015) and Brussels (2016) incidents has prompted the revision of counterterrorism strategies and strengthened police cooperation within the European Union. The United Nations has also enshrined, through its resolutions, the obligation of states to adopt legislative, administrative, and operational measures for the prevention and suppression of terrorism. The European Union has transposed these standards through legal instruments, operational mechanisms, and specialized structures, while Romania has aligned its domestic framework, institutional responsibilities, and investigative techniques with current risks [3].

At the national level, counterterrorism efforts are closely linked to broader initiatives aimed at preventing and combating organized crime, given the increasingly evident intersection between the two phenomena particularly in areas such as financing, arms trafficking, use of forged documents, and logistical support. Terrorism-related investigations require a high degree of interinstitutional cooperation, operational coordination, and procedural standards compatible with those of the European Union.

In this context, the present study aims to:

- Highlight the essential legislative landmarks (UN, EU, Romania).
- Analyze the dynamics of the terrorist phenomenon and its implications for policing activities.
- Present modern operational directions, including elements of prevention, early detection, intervention, investigation, and international cooperation.

The article approaches terrorism from a policing perspective, emphasizing an integrated operational cycle prevention, counteraction, and investigation as well as the need to optimize both punitive and non-punitive instruments while safeguarding fundamental rights. Through this structure, the research combines legal analysis with contemporary operational methodologies, offering a realistic and applied perspective on counterterrorism.

2. Legal Frameworks in Counterterrorism

The fight against terrorism is grounded in a comprehensive set of legal instruments operating across three levels: global (United Nations), regional (European Union), and national (Romania). These instruments establish state obligations, minimum action standards, and key principles governing police cooperation, information exchange, prevention of radicalization, and the conduct of specific operational activities.

2.1. United Nations Level - Global Standards

The United Nations has gradually developed a universal legal framework founded on the obligation of states to prevent, investigate, and punish terrorist acts. The core documents include:

UN Security Council Resolution 1373 (2001), adopted in the aftermath of the September 11 attacks, which requires states to: criminalize terrorist-related acts; freeze terrorist funds and financial resources; strengthen international cooperation (extradition, mutual legal assistance, data exchange); reinforce border control mechanisms.

The United Nations Global Counter-Terrorism Strategy (2006, revised 2010-2016), which introduced a proactive and police-preventive approach, emphasizing cooperation among national agencies, measures against radicalization, and the reduction of factors conducive to terrorism.

A series of UN sectoral conventions, addressing aircraft hijacking, protection of diplomats, terrorist financing, and bomb attacks, among others, which oblige states to criminalize such acts and cooperate internationally.

Key UN principle: Terrorism is a global threat that requires a coordinated response focused on prevention, cooperation, and disruption of financing channels [4].

2.2. European Union Level - Common Security and Police Cooperation

The European Union has developed one of the most advanced regional systems for combating terrorism, combining criminal legislation with operational mechanisms and intelligence-based instruments.

Key documents and mechanisms:

The EU Counter-Terrorism Strategy (Council of the EU, 2005), structured around four pillars: *Prevent - Protect - Pursue - Respond*.

Directive (EU) 2017/541 on Combating Terrorism, which obliges Member States to criminalize:

- preparation and facilitation of terrorist acts.
- recruitment and training for terrorism.
- travel for terrorist purposes ("foreign fighters").
- terrorist financing.

Operational instruments:

- *Europol* - European Counter Terrorism Centre (ECTC).
- *Eurojust* - judicial coordination.
- *SIS II*, *PNR*, and *Prüm* systems - data exchange for suspect identification.
- *Frontex* - border management and detection of terrorist routes.

The EU model is operational in nature, emphasizing real-time police and intelligence cooperation within a common security framework.

2.3. Romania's Level - National Legislative and Institutional Framework

Romania has aligned its legislative and operational systems with UN and EU standards.

Core elements:

Law No. 535/2004 on the Prevention and Combating of Terrorism, the main legislative act, which:

- defines terrorism and related offenses.
- establishes the *National System for Preventing and Combating Terrorism (SNPCT)*.
- designates the Romanian Intelligence Service (SRI) as the national authority in this field.
- regulates interinstitutional cooperation and operational measures.

Key institutions:

- *SRI* - coordination of SNPCT and intelligence component.
- *Ministry of Internal Affairs (MAI)* - Romanian Police, Gendarmerie, SIAS units, and other operational structures.
- *DIICOT* - criminal prosecution in terrorism-related cases.
- *MAE, MApN, SPP, STS* - specialized institutional support.

Complementary legislation:

- The *Criminal Code* and *Criminal Procedure Code* (special investigative measures).
- The *Law on the Prevention and Combating of Terrorist Financing*.
- Regulations on border control, population records, and identity documents.

Characteristic of the Romanian system: a hybrid model combining a strong intelligence component (SRI) with an operational policing one (MAI + DIICOT).

The three levels UN, EU, and Romania together constitute a coherent legislative framework. However, actual effectiveness depends on the operational dimension, which will be examined in subsequent chapters: prevention, investigation, cooperation, operational analysis, and the application of modern counterterrorism tools [5].

3. The Dynamics of the Terrorist Phenomenon and the Operational Dimension in the Digital Environment

For practitioners, terrorism is not an abstract category but a process with distinct phases - recruitment, financing, training, mobilization, execution, consolidation/propaganda - and each phase requires specific detection and intervention tools and methods. This section outlines the dynamic elements that must be monitored at the operational level.

Contemporary terrorism can no longer be understood outside the virtual space. Whereas in the past terrorist organizations relied primarily on physical contacts, clandestine networks and traditional logistical circuits, the digital environment now provides a complete ecosystem for recruitment, propaganda, financing, encrypted communication, procurement, operational planning and training. Cyberspace thus becomes the force multiplier of modern terrorism, and the interaction between terrorism and cybercrime produces a hybrid threat with unprecedented dynamics.

3.1. General Considerations from a Policing Perspective

A policing approach to counterterrorism requires the integration of the normative framework with proactive operational mechanisms oriented toward prevention, monitoring, deterrence and the neutralization of radicalized groups or individuals. In the current security climate marked by intensified online radicalization, increased cross-border mobility and ready access to highly lethal means of attack police authorities must adopt flexible, integrated methods founded on real-time information sharing.

In Romania, operational responsibilities are distributed among specialised structures of the Ministry of Internal Affairs (MAI), the Romanian Intelligence Service (SRI), the Public Prosecutor's Office and European cooperation bodies. Specifically, the police play a central role in: early identification of risk indicators, collection and exploitation of operational information, documentation of offences, executing procedural actions and supporting anti-terrorist intervention units during the critical phase of a threat. This approach reflects current counterterrorism principles, whereby prevention remains the primary strategic objective and the use of force is a last resort, applicable only when all preventive mechanisms have failed.

Police investigations into terrorist modus operandi focus on the operational chain: radicalization → financing → recruitment → planning → logistic preparation → execution → claim of responsibility → propaganda and regeneration. Each link in this chain represents an opportunity for intervention and disruption, which explains why modern strategies emphasize early detection of intent and discreet surveillance of high-risk environments (unauthorised places of worship, clubs, prisons, encrypted forums, social networks, etc.).

Consequently, the policing perspective is characterised by several defining features:

- orientation toward prevention and anticipation.
- action based on operational intelligence rather than merely post-factum reaction.
- a multi-institutional approach.
- continuous surveillance of risk environments and cross-border flows.
- rapid and lawful intervention during the early stages of attack preparation.

This operational logic is consistently reflected in European practice, and the attacks in Paris (Bataclan, 2015) and Brussels (2016) demonstrated that the window between radicalization and action can compress dramatically, sometimes to only a few days, thereby imposing accelerated investigations and efficient police-cooperation mechanisms.

3.2. The Role of Operational Structures and Inter-Institutional Cooperation

Countering terrorism requires a complex institutional architecture in which policing structures operate in an integrated manner with intelligence services, specialised public prosecutors' offices and international organisations. From the perspective of operational investigation, the police have two fundamental missions: the acquisition of information and the disruption of a potential perpetrator's operational capacity. This role can be fulfilled only within a cooperative framework in which data exchange, integrated analysis and the synchronization of interventions are functional.

Nationally, cooperation is principally effected through coordination structures and joint operational centres representing the Ministry of Internal Affairs, competent prosecution offices and the Romanian Intelligence Service. The dynamics of the phenomenon oblige institutions to communicate in real time, because the operational window for preventing an attack is narrow and any bureaucratic bottleneck can transform a mere operational signal into a tragedy. The Bataclan explosion in 2015 and the Brussels attacks in 2016 demonstrated at the European level that the absence of rapid information exchange among law-enforcement bodies can allow terrorist cells to travel, finance themselves and equip without detection.

At the international level, police forces use an extensive cooperative toolkit that includes:

- *Europol* (EC3, ECTC) - operational analysis, support in investigating networks, cross-border forensic intelligence.
- *Interpol* - databases on wanted persons, forged documents, weapons and foreign terrorist fighters (FTF).
- *SIRENE & SIS II* - real-time alert flows for suspects and fraudulently used documents.
- *Frontex* - a critical actor in mobility and border control for filtering risks.

Such cooperation enables national police to see beyond their borders, which is essential since contemporary terrorism does not respect territorial limits.

3.3. Radicalization, Propaganda, Financing and Recruitment in the Online Environment

Social platforms, forums and encrypted messaging applications have become the principal instruments through which terrorist networks:

- disseminate highly emotive visual and narrative propaganda.
- recruit vulnerable youth within the EU.
- create ideological bubbles that are difficult for authorities to penetrate.
- organize micro-digital communities that later translate into offline action.

ISIS, Al-Qaeda and their affiliates have developed sophisticated media factories, exploiting Twitter, Telegram, TikTok, Discord and the digital diaspora. Radicalization is no longer centralized but individualized and automated, assisted by platform algorithms that promote extremist content to susceptible users.

The Dark Web - the hidden infrastructure of digital terrorism

The Dark Web functions as a global marketplace for illegal goods and services, including:

- weapons and components for improvised explosive devices (IEDs).
- personal data, bank accounts and stolen identities.
- hacking services, ransomware or DDoS capabilities.
- false passports and transit documents.
- manuals on explosive-making, urban tactics and counter-surveillance techniques.

This parallel space allows terrorists to purchase goods, communicate and obscure their traces thanks to:

- TOR / I2P networks.
- PGP encryption.
- escrow platforms.
- bulletproof servers hosted in hostile jurisdictions.

Cryptocurrency financing of terrorism has, in recent years, moved beyond isolated experiments to become a relevant component of the digital underground economy associated with jihadist organisations and other extremist groups. Terrorists have adopted virtual currencies as a funding instrument that is difficult to trace:

- Bitcoin, Monero, Zcash.
- anonymised electronic wallets.
- mixers/tumblers for laundering funds.
- covert fundraising through sham NGOs or crypto-charity campaigns.

Operational Advantages for Terrorist Networks

Advantage (English)	Explanation
Relative anonymity	Conceals the true identity of the ultimate beneficiary.
Cross-border nature	Operates beyond the scope of conventional banking controls.
Speed	Transfers can be executed within minutes.
Convertibility to cash	Can be exchanged for cash via cryptocurrency ATMs or informal exchangers.

These mechanisms complement traditional funding sources (trafficking, fraud, donations) rather than replacing them, rendering the phenomenon more resilient and less conspicuous.

According to Europol and the Financial Action Task Force (FATF), the current trend is no longer the concentration of large sums in a single transaction but the fragmentation into micro-flows below the detection thresholds of banking systems - a method perfectly compatible with autonomous terrorist cells and low-cost attacks.

Encrypted Communication and Cooperation between Terrorists and Hackers

Terrorist actors employ end-to-end encryption (e.g., Telegram, Signal, ProtonMail) to evade interception. Increasing evidence indicates that certain terrorist groups:

- collaborate with financially motivated hackers (crimeware-as-a-service).
- purchase exploits and malware.
- commission cyberattacks against critical infrastructure.

Thus, the relationship between terrorism and cybercrime is evolving from a mere connection into a complementary pact.

3.4. The Operational Cycle of Counterterrorism Investigation

Counterterrorism investigations follow a specific cycle distinct from classical criminal inquiries, as their main objective is not to prove the offence after it has occurred, but to interrupt the operational chain before the lethal outcome materializes. In both doctrine and practice, the police counterterrorism cycle includes:

- Identification of early warning signals - monitoring high-risk environments, online surveillance, behavioural analysis, and assessment of interactions with potential radical structures.
- Collection and verification of information - HUMINT, SIGINT, OSINT, technical surveillance, and undercover investigations.
- Assessment of threat level and target prioritization - according to intent, capability, and proximity to action.
- Documentation and evidentiary consolidation - as far as possible without compromising intelligence operations.
- Neutralization through coercive measures - arrest, searches, logistical disruption, and freezing of financial flows.
- Post-intervention intelligence exploitation - aimed at preventing cell regeneration.

This cycle implies that, operationally, the police act with one hand in the intelligence domain and the other in the criminal justice sphere, in finely tuned synchronization with intelligence services and specialized prosecutors' offices.

3.5. Operational Prevention and Counter-Radicalization

Prevention constitutes the first line of defence in counterterrorism, grounded in the concept of *"intervention before intent"* - identifying radicalization processes early and disrupting them before they evolve into violent action.

From a policing standpoint, prevention has two complementary dimensions:

- Operational (hard) prevention - surveillance of individuals and environments with terrorist potential, monitoring vulnerable areas, gathering intelligence, and conducting deterrence actions.
- Community and social (soft) prevention - local partnerships with relevant social actors (schools, families, religious institutions, NGOs, and local authorities).

The radicalization process is gradual and long-term, fueled by socio-economic frustrations, ideological narratives, and digital influences.

Operationally, the police have a duty to understand high-risk environments, engage with them, and intervene proactively through:

- identification and limitation of the influence of radical "preachers".
- monitoring of spaces where violent propaganda circulates.
- collaboration with community leaders and social services.
- early notification of intelligence agencies when clear signs of radicalization emerge.

The online environment currently represents the most dynamic arena for radicalization, as it offers anonymity, access to extremist materials, direct contact with recruiters, and the creation of "ideological bubbles."

Consequently, modern prevention cannot exist without a strong *cyber-police* and *cyber-intelligence* component.

3.6. Combating the Financing of Terrorism - A Policing Perspective

Financing represents the *"oxygen of terrorist organizations."* Even small cells operating under a low-cost model as seen in numerous European attacks between 2015 and 2018 require funding for:

- logistical expenses (safe-house rent, transport, communication).
- procurement of weapons or chemical precursors.
- forgeries, documentation, and operational support.

The role of the police is to asphyxiate financial resources, using mechanisms that include:

- tracing banking flows through Financial Intelligence Units (FIUs).
- monitoring suspicious transactions, including micro-payments.
- cooperating with private financial institutions (banks, non-banking entities).
- tracking online funding through cryptocurrencies, crowdfunding, or disguised donations.
- applying extended confiscation and rapid precautionary measures before funds are dispersed.

European experience shows that contemporary jihadism relies on a hybrid mix of sources: personal savings, cyber fraud, trafficking, "charitable" donations, cryptocurrencies, and funding from state sponsors. Therefore, police authorities must treat the financial component as a mandatory parallel investigation, not as a secondary element [6].

3.7. Modern Operational Techniques in Countering Terrorism

The counterterrorism operational field has shifted into a hybrid zone in which police must combine classical methods with intelligence tools and technological expertise. The most commonly employed modern techniques are summarized below.

Category	Operational Tools
SURVEILLANCE	physical tailing, operational surveillance techniques, geolocation, drones in urban environments
OSINT	social-media monitoring, encrypted forums, Dark Web monitoring
SIGINT	interception, communications analysis, metadata analysis
Undercover HUMINT	infiltrators, collaborators, human sources within the diaspora or religious communities
Digital forensics	exploitation of phones, servers, cloud services, and crypto-wallets

Compared with organised-crime investigations, counterterrorism inquiries demand greater speed because the optimal intervention moment is prior to the execution phase; police must therefore apply a zero-tolerance approach to risk. This requires a high operational standard, including:

- parallel documentation of evidence for the Prosecutor's Office.
- filtering of raw intelligence through analytical units.
- "hit-and-freeze" actions: entry - capture - neutralization.

All these techniques are reflected in national doctrine, which emphasizes early intervention, disruption of logistics, and separation of the perpetrator from their resources.

3.8. Tactical Intervention Management and the Neutralization of Terrorist Cells

Neutralizing a terrorist actor requires an integrated operational chain that combines the intelligence component with tactical action. Unlike organised crime - where intervention timing may be chosen strategically to document an entire network - in terrorism the priority is not necessarily total network dismantlement but the prevention of loss of life.

Modern police and counterterrorism procedures for tactical intervention are based on the following principles:

Intervention within the "short window" - act immediately once there are clear indications that an attack is imminent. Waiting is excluded, as it risks irreversible consequences.

- Segmentation of the operational area - rapidly separate suspects from civilians, isolate the perimeter, and control flows (evacuation, communications, transport).
- Neutralization with proportionate force - apply force progressively but with a clear mandate: the suspect must be neutralized before they can detonate, attack, or escape.
- Protection of evidence and the information chain - to ensure judicial finality, electronic, biological and documentary evidence must be preserved and processed immediately.
- Priority on "capture" rather than "elimination" when feasible - to enable intelligence exploitation, particularly regarding the logistics network (financing, recruitment, communications).

European experience (Belgium 2016, Germany 2018, France 2020) has shown that mixed teams (intervention - negotiation - technical exploitation - analysts) increase operational efficiency by over 50%, as they provide a multi-dimensional real-time response [7].

3.9. International Cooperation and Operative Information Exchange

Terrorism is transnational in terms of mobility, financing, communications and ideology; consequently, a purely national response is structurally insufficient. Modern policing operates within a cooperative architecture that includes:

- *Europol* (ECTC - European Counter-Terrorism Centre) - an analytical and operational hub for terrorist situations with real-time database access.

- *Interpol* - international notices and follow-ups, including those concerning Foreign Terrorist Fighters.
- *SIRENE* and *SIS II / SIS Recast* - immediate operational exchange within the EU.
- *PNR, API, FRONTEX* - for monitoring passenger and border flows.
- informal HUMINT networks obtained via internal-affairs attachés, liaison officers and Joint Investigation Teams (JITs).

A critical element in police counterterrorism work is the speed of data exchange. In the terrorism domain, "delayed" information is effectively useless. Therefore, the current European architecture is built on:

- exchange in minutes and hours, not days and weeks.
- direct access to shared databases without bureaucratic detours.
- interoperability of databases (SIS, VIS, Eurodac, PNR, ECRIS).

These mechanisms are particularly effective in identifying fighters returning from Syria-Iraq, mobile propagandists and cells traversing Schengen routes.

Countering terrorism from a policing perspective entails:

- early prevention through monitoring, social education and counter-radicalization.
- disruption of violent actors' logistical and financial resources.
- rapid tactical intervention aimed at saving lives.
- continuous investigation and intelligence exploitation even after the target has been neutralized.
- permanent international cooperation, in the logic of "zero gaps between systems."

If organised crime pursues profit, terrorism pursues impact; the only effective response is proactive, integrated and decisive policing, supported by law but not constrained by it operationally. In this equation, time becomes the primary operational currency: whoever acts first wins [8].

4. General Conclusions and Modern Directions in Counterterrorism Action

4.1. General Considerations

Contemporary terrorism is no longer an isolated manifestation but a complex form of organized violence encompassing political, ideological, religious, and technological dimensions. In the 21st century, it continuously reinvents itself, adapting to geopolitical shifts, technological vulnerabilities, and the social tensions of a globalized world.

From a policing perspective, the fight against terrorism extends far beyond identifying and neutralizing perpetrators. It involves the integrated management of the broader criminal ecosystem surrounding the phenomenon of radicalization, propaganda, financing, logistical support, recruitment, and the exploitation of the digital space.

Modern police forces should be regarded as *operational intelligence actors*, not merely reactive forces. They occupy the center of the preventive, investigative, and intervention mechanism, serving as the connective node between the legal, intelligence, and tactical domains within a unified framework of coordinated action.

4.2. Integrating the Legislative Dimension into the Policing Approach

The legal framework-national, European, and international-provides both the foundation of legality and the operational flexibility required by field structures. The UN counterterrorism conventions, the EU Framework Decisions and Directives, and Romanian legislation (Law No. 535/2004 on the Prevention and Combating of Terrorism, the Criminal Code, and the National Security Law) form a coherent system designed to:

- legally define the terrorist phenomenon.
- establish the competences of responsible institutions.

- regulate special investigative and intervention means.
- ensure sanctions proportionate to the gravity of the acts committed.

However, the effectiveness of these instruments depends not merely on their normative existence but on the *operational capacity and interinstitutional coordination* behind their implementation. Where the law provides the framework, the police must provide the action.

4.3. Modern Directions in Counterterrorism Action

Recent European and global developments reveal five key modern directions in counterterrorism, particularly relevant for police structures:

- Artificial Intelligence and Predictive Analysis - employing algorithms for detecting suspicious behavior, facial recognition, and correlating multi-source data flows (SIS, Europol, Interpol, OSINT).
- Countering Online Radicalization - monitoring extremist narratives, removing radical content, and developing counter-messaging through educational and civic platforms.
- Strengthening International Cooperation - enhancing database interoperability, creating regional analysis and intervention centers, and ensuring real-time sharing of operational intelligence.
- Adapting Urban Intervention Tactics - training police units for rapid response in crowded environments and neutralizing mobile targets while respecting the principle of proportionality.
- Combating Terrorist Financing and Logistical Support - reinforcing Financial Intelligence Units (FIUs), tracing cryptographic transactions, and cooperating with private-sector entities (banks, fintech, transport, communications).

These directions reflect the increasing integration of technology and behavioral analysis into police operations, marking the transition from *reactive policing* to *anticipatory policing*.

4.4. Recommendations for Policing Practice

Based on legislative, operational, and institutional analysis, the following recommendations can be formulated:

- Develop data-analysis capabilities (*big data policing*) and train personnel in OSINT and cyber-intelligence.
- Strengthen the culture of partnership among police, community, and the private sector.
- Implement unified response mechanisms (*joint task forces*) for terrorist-risk situations.
- Establish a National Counterterrorism Training Centre within the Ministry of Internal Affairs, modeled after Europol's ECTC.
- Intensify the exchange of best practices with EU member states and NATO partners in the fields of *tactical intelligence* and *counter-radicalization*.

4.5. Final Conclusions

Terrorism cannot be completely eradicated, but it can be controlled and prevented through anticipation, coordination, and decisive action.

The modern state must base its response on two complementary pillars: law and operational capability. Within this framework, the police is no longer merely an enforcer of the law, but a strategic actor of national and international security. Its capacity to collaborate, analyze, and respond swiftly represents the most effective deterrent against terrorism.

The prevailing consensus in the specialized literature is unequivocal: the next principal arena of terrorist financing will be the crypto-digital environment. Consequently, the fight can no longer rely solely on classical legal instruments but must incorporate cyber-investigation, transnational cooperation, and digital policing techniques.

References

[1]. Security Council of UN, 3 DECEMBER 2001, “Resolution 1373”, Available: https://legislatie.just.ro/Public/DetaliiDocument/32477.

[2]. Ștefan-Gabriel DASCĂLU, 2016 *“Combaterea criminalității organizate”* publisher: SITECH, city: Craiova, pp. 335-339.

[3]. United Nations, 2006, “Global Counter-Terrorism Strategy.” Available: https://www.un.org/counterterrorism/un-global-counter-terrorism-strategy.

[4]. European Council, 2018, “UE Strategy for countering terrorism”, Available: https://eur-lex.europa.eu/TodayOJ/.

[5]. EUROPOL, 2016-2017 “Terrorism Situation and Trend Report”, Available: https://www.europol.europa.eu/publications-events/main-reports/tesat-report.

[6]. OSCE, 25 March 2019, “Preventing Terrorist Radicalization”, Available: https://www.osce.org/event/osce-wide-counter-terrorism-conference-2019.

[7]. Financial Action Task Force, 2022 “FATF Annual Report 2022-2023”, Available: https://www.fatf-gafi.org/en/publications/Fatfgeneral/FATF-Annual-report-2022-2023.html.

[8]. EUROPOL, 2022 “European Counter Terrorism Centre - ECTC operational report”, Available: https://www.europol.europa.eu/about-europol/european-counter-terrorism-centre-ectc.

AI-Assisted Anomaly Detection for Cybersecurity in IMS Core Networks: A KPI-Driven Study Based on Real-World Telecom Data

Bianca-Ștefania VĂDUVA
Faculty of Electronics, Telecommunications and Information Technology,
National University of Science and Technology POLITEHNICA Bucharest, Romania
bianca.vaduva@stud.etti.upb.ro

Abstract

In modern IP Multimedia Subsystem (IMS) core networks, the detection and prevention of cybersecurity threats remain a critical challenge due to the dynamic nature of signaling traffic and the increasing complexity of infrastructure. This paper proposes an AI-assisted anomaly detection approach based on statistical modeling of key performance indicators (KPIs) collected from real-world telecom networks over a one-month period. The analysis targets multiple IMS elements across two major network regions, focusing on Call Setup Success Rate and Total Traffic (Erlang). A contextual z-score model was implemented in MATLAB to monitor these KPIs per hour, enabling the identification of time-based deviations without relying on static thresholds. An alert logic was added to mark days with excessive anomaly rates (>5%) as potentially suspicious. A major traffic spike detected on March 1st is analyzed as a case study, suggesting a possible signaling flood or operational event. The results demonstrate the feasibility of unsupervised anomaly detection in IMS environments, providing early warning signals for cybersecurity-related incidents. This KPI-driven methodology can be extended with advanced AI models for predictive alerting and integration with network management systems.

Index terms: anomaly detection, artificial intelligence, cybersecurity, IMS core networks, KPI monitoring

1. Introduction

The IP Multimedia Subsystem (IMS) has become a fundamental architecture in modern telecommunications, enabling the delivery of multimedia services such as voice, video, and messaging over IP-based networks. As IMS deployments scale to support a growing number of users and services, their exposure to cybersecurity threats has increased significantly [1], [2]. The complexity of signaling protocols like SIP and Diameter makes IMS infrastructures particularly vulnerable to attacks such as signaling floods, malformed messages, and denial-of-service (DoS) attempts [1], [3].

Conventional security mechanisms in telecom environments - such as firewalls, static filters, or signature-based intrusion detection - are often ineffective against dynamic or unknown threats. Even minor deviations in signaling behavior can indicate the early stages of a cyberattack, especially when they impact critical components such as Session Border Controllers (SBCs) or Call Session Control Functions (CSCFs), which are responsible for managing and securing signaling flows [2], [4].

These challenges are further amplified in next-generation telecom environments, where convergence between fixed and mobile networks, together with the integration of IoT and post-quantum resilience concerns, increases the overall complexity and risk surface [5], [6], [7]. In such contexts, traditional perimeter defense becomes insufficient, and monitoring internal KPI behavior becomes a more adaptive way to detect early-stage anomalies.

This work proposes an AI-assisted anomaly detection framework tailored for IMS core networks. The system operates on real-world telecom KPI data collected over a one-month period and focuses on detecting deviations in Call Setup Success Rate and Total Traffic (Erlang). By applying contextual statistical modeling, the framework can identify operational anomalies that may correspond to cyber threats or hidden configuration issues.

The main contributions of this work are: (1) the design and implementation of an unsupervised anomaly detection model adapted to IMS traffic behavior, (2) the use of real operational data from multiple regions to validate the model, and (3) a discussion on how such systems can enhance cybersecurity readiness in telecom infrastructures.

The proposed framework can be extended with more advanced AI models and integrated into existing network management or security information and event management (SIEM) systems for real-time threat detection.

2. Cybersecurity Challenges in IMS

2.1. Signaling Protocol Vulnerabilities

The IMS architecture relies heavily on signaling protocols such as SIP and Diameter, which were not originally designed with strong built-in security mechanisms. As a result, these protocols are vulnerable to malformed messages, spoofing, and denial-of-service (DoS) attacks. SIP floods using REGISTER or INVITE methods can overload Call Session Control Functions (CSCFs), leading to service unavailability and control-plane saturation. This type of vulnerability is especially concerning in legacy PSTN environments that are transitioning to IMS architectures [1], [3]. In addition, attackers may exploit optional or poorly validated SIP headers to bypass filters or cause parsing failures.

2.2. Weaknesses in Implementation and Configuration

Although the IMS standard includes security mechanisms such as IPSec, TLS, and message integrity protection, their implementation is often inconsistent across operational networks. In practice, encryption is frequently limited to access segments, while core or inter-domain links - such as those between SBC and CSCF nodes - may remain unprotected due to performance trade-offs or interoperability constraints [2]. Furthermore, misconfigurations related to DNS resolution, HSS access control, or SIP session timers can expose new vulnerabilities or amplify the impact of otherwise benign anomalies [7].

2.3. Limitations of Traditional Security Mechanisms

Traditional telecom security mechanisms - such as access control lists, static firewalls, and signature-based intrusion detection systems - are generally inadequate for handling evolving and dynamic threat patterns. These tools rely on predefined rules and known signatures, leaving them unable to detect novel or slow-paced attacks. In IMS environments, where even slight deviations in signaling behavior can cascade into large-scale service degradation, this limitation is especially critical [4].

2.4. The Role and Limitations of SBCs

Session Border Controllers (SBCs) play a central role in securing IMS networks, acting as gatekeepers for signaling and media traffic. Their responsibilities include protocol validation, topology hiding, DoS protection, and interconnect policing. However, under high-load or attack conditions, SBCs themselves can become bottlenecks. When operating with static rule sets or limited adaptability, their ability to detect and mitigate emerging threats is significantly reduced [4].

2.5. The Case for Anomaly Detection

To address the limitations of conventional defense mechanisms, modern IMS security architectures increasingly integrate anomaly detection techniques. These approaches monitor real-time traffic behavior, compare it against learned baselines, and flag deviations - rather than relying solely on known attack signatures. When applied to key performance indicators (KPIs) such as Call Setup Success Rate and signaling traffic volume, anomaly detection can help surface otherwise hidden or emerging threats. This capability is particularly valuable in IMS environments, where latency-sensitive services require proactive identification of degradation events [8].

3. Methodology

a. Data Collection

To ensure a practical and representative foundation for anomaly detection in IMS environments, this study utilizes real operational data gathered from a commercial telecom network over a one-month period (February 27 to March 27, 2025). The dataset includes measurements from multiple Session Border Controller (SBC) nodes located in two distinct network regions: Brașov and Cluj. SBCs serve as strategic control points at the edge of the IMS core, where they enforce signaling security, manage routing policies, and regulate access control [5].

The data consists of hourly samples exported from performance monitoring systems, structured across multiple Excel sheets-each corresponding to a specific SBC. Two primary key performance indicators (KPIs) were selected: **Call Setup Success Rate**, representing the ratio of successful call setups to total attempts, and **Total Traffic (Erlang)**, a volumetric measure of signaling load processed by each SBC.

These metrics were chosen to enable both temporal and spatial analysis of network health and to facilitate regional comparisons and node-specific anomaly detection.

b. Feature Selection

The KPIs used in this study were selected based on their operational relevance and direct relationship with signaling behavior and service availability. Call setup failures, for instance, can indicate issues such as signaling errors, registration flooding, or control-plane congestion [2]. Similarly, sudden spikes in traffic may result from replay attacks, malformed signaling bursts, or misconfigured user agents [4].

All features were extracted with an hourly granularity to strike a balance between resolution and noise tolerance. Each KPI stream was treated independently per SBC, with no prior labeling, making the detection process fully unsupervised.

c. Detection Algorithm

The anomaly detection model is based on contextual z-score deviation, calculated independently for each hour of the day. This approach accounts for normal daily variations (e.g., low traffic at night, high peaks in the afternoon) and is robust to seasonality in telecom workloads.

For each data point $x_{i,h}$, where h is the hour, the z-score was computed as:

$$z_{i,h} = \frac{x_{i,h} - \mu_h}{\sigma_h} \tag{1}$$

where μ_h and σ_h are the mean and standard deviation for all data points at hour h. A threshold of $|z| > 2.5$ was used to flag anomalies, which is commonly applied in statistical process control and time series anomaly detection [5].

d. Alert Logic

Beyond individual outlier detection, the system implements an aggregated alerting mechanism at the daily level. For each day, the total number of anomalies is divided by the total number of hourly samples, yielding a daily anomaly rate. Days exceeding a 5% threshold are flagged as suspicious.

This aggregation helps reduce operator overload and aligns with anomaly clustering practices found in modern cybersecurity telemetry systems [8]. Similar logic is used in operational environments where lightweight alerting systems monitor deviations from baseline across multiple nodes and metrics [6].

e. Implementation

The detection framework was implemented in MATLAB, leveraging its high-performance computing and data visualization capabilities. Built-in statistical functions and table structures were used to preprocess KPI streams, compute z-scores, and generate visual outputs.

Each SBC's time-series behavior was visualized individually, with red markers indicating detected anomalies. Additionally, a heatmap was constructed to provide an at-a-glance view of daily alert status across all SBCs. The implementation is modular and extensible, allowing for the inclusion of new KPIs or streaming data pipelines. The unsupervised approach ensures applicability across heterogeneous IMS environments and vendor technologies.

4. Experimental Results

The proposed anomaly detection framework was applied to real operational KPI data from six SBC nodes - three located in the Brașov region (brasbc01, brasbc02, brasbc03) and three in Cluj (clusbc01, clusbc02, clusbc03). The monitored KPIs included Call Setup Success Rate and Total Traffic (Erlang), analyzed independently per node over a one-month period (February 27 - March 27, 2025).

4.1. Detected Anomalies per SBC

Table 1 presents the number of data points analyzed and the anomalies detected per SBC. For the Call Setup Success Rate, most anomalies were detected during off-peak hours - particularly at night and during weekends - when normal variation is more pronounced. In contrast, anomalies in Total Traffic (Erlang) were rare, with only one extreme outlier observed.

Table 1. Detected anomalies per SBC node

SBC Node	Total Points	Anomalies Detected
brasbc01	2689	110
brasbc02	2689	98
brasbc03	2689	94
clusbc01	2689	91
clusbc02	2689	99
clusbc03	2689	118

In addition to the numerical summary in Table 1, the anomaly detection process was visualized for each SBC using a time-series plot with overlayed anomaly markers. Figure 1 illustrates a representative example, where the KPI values are plotted over time, and detected anomalies are highlighted in red.

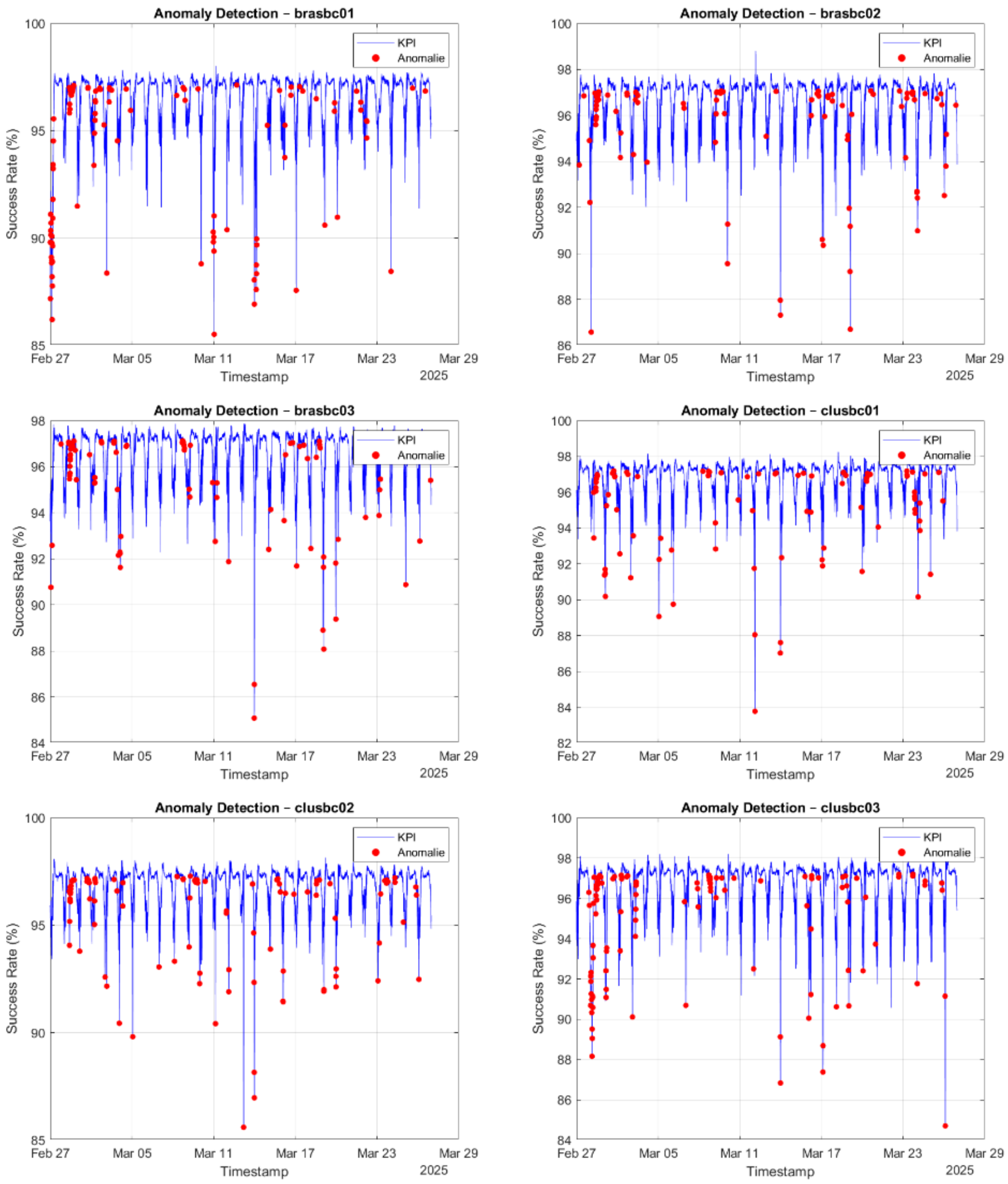


Fig. 1. Time-series anomaly detection for Call Setup Success Rate - sample SBC

4.2. Daily Alert Heatmap

Beyond individual anomaly detection, an aggregated daily alerting mechanism was implemented. For each day, the proportion of anomalous KPI samples was computed, and days exceeding a 5% anomaly threshold were flagged as suspicious. This higher-level abstraction helps reduce noise from isolated outliers and directs attention toward periods of potential operational or security relevance.

Figure 2 presents a heatmap summarizing the alert status for each SBC across the monitored period. Most days were categorized as normal, indicating stable network behavior. However, several isolated alert days were identified, suggesting short-term anomalies, localized performance degradations, or possible low-volume signaling threats.

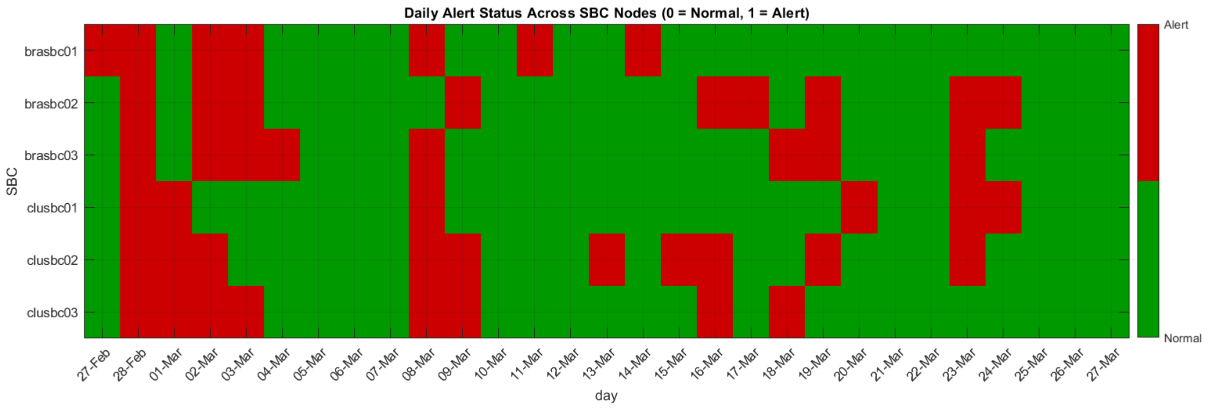


Fig. 2. Daily anomaly alert heatmap (Call Setup Success Rate)

4.3. Spike Detected on March 1st

A significant anomaly was observed on March 1st, characterized by an abrupt spike in Total Traffic (Erlang). The recorded traffic volume on this day was drastically higher than the contextual hourly average, resulting in a z-score exceeding the anomaly threshold by a wide margin. Figure 3 clearly shows this outlier compared to the typical traffic profile for the node.

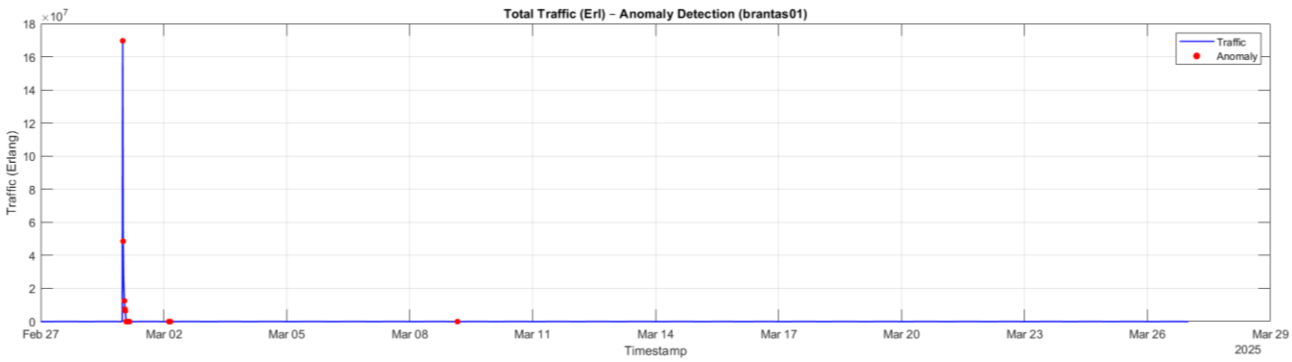


Fig. 3. Anomaly detection - Total Traffic on brantas01

This spike may be explained by an operational event such as rerouted traffic following a failover, or alternatively, a deliberate volumetric attack (e.g., signaling flood). While no correlated degradation in other KPIs was observed, such singular deviations justify investigation and support the utility of anomaly-based early detection strategies [8].

4.4. Comparative Regional Analysis

The comparative analysis between the two regions revealed that Cluj SBCs exhibited a slightly higher anomaly density in the Call Setup Success Rate KPI. This may reflect differences in network configuration, traffic distribution, or signaling load per node. Brașov SBCs displayed greater overall stability, with the exception of the isolated high-traffic anomaly observed on March 1st. These differences underscore the value of localized monitoring, enabling telecom operators to detect and respond to region-specific risks or misconfigurations.

5. Discussion

The experimental results confirm that KPI-driven anomaly detection can serve as an effective early warning system for both cybersecurity incidents and operational issues within IMS core networks. The use of contextual statistical modeling allows for the identification of irregular patterns that might remain undetected by static or rule-based monitoring systems.

a. Interpreting the March 1st Event

The most significant anomaly observed during the analysis period was a pronounced traffic spike on March 1st in the brasbc01 node. This deviation in the Total Traffic (Erlang) KPI is highly unlikely to be the result of normal user behavior. Potential explanations include: traffic rerouting due to failover mechanisms in other network segments, automated bulk provisioning processes or the initiation of a deliberate signaling flood as part of a denial-of-service (DoS) attempt.

Although no concurrent degradation in other KPIs (such as Call Setup Success Rate) was recorded, such isolated traffic surges may indicate early-stage volumetric attacks, designed to probe system resilience without triggering service collapse. Traditional intrusion detection systems (IDS) are often unable to detect such events due to their low signature visibility, but statistical models - such as the one implemented here - can surface them through behavioral deviation [4].

b. Operational Relevance of Detected Anomalies

The anomalies detected in Call Setup Success Rate were predominantly concentrated during off-peak hours, especially at night and on weekends. These may align with routine network maintenance activities or scheduled reconfigurations. However, repeated minor anomalies under light traffic conditions may also reveal: improper session timer configurations, unsolicited external signaling probes or underlying control-plane fragility.

Rather than filtering such events out, the system's ability to detect both acute and low-severity anomalies allows for a layered observability strategy, blending performance monitoring with cybersecurity insight [8].

c. Integration with Existing Telecom Monitoring Systems

A major advantage of the proposed framework is its lightweight, unsupervised architecture, which enables easy integration into existing telecom operations. Since the system does not require labeled training data, it can function alongside traditional monitoring solutions such as NOC dashboards or SIEM platforms.

SBCs and CSCFs already generate extensive telemetry, including detailed KPI logs. These data streams can be used as input for real-time anomaly detection modules. Once deployed, the contextual z-score model can enhance automated alerting workflows and reduce time-to-detection for emerging threats.

d. Limitations and Future Considerations

The current implementation focuses on univariate anomaly detection, analyzing each KPI independently. While effective, this approach does not capture inter-metric correlations. Future work may expand into multivariate anomaly detection, incorporating features like call failure types, signaling latency, and error rates.

Another potential extension involves anomaly classification, distinguishing between types of deviations (e.g., spikes, drops, persistence) to improve alert prioritization.

As more advanced AI techniques become accessible, methods such as isolation forests, autoencoders, or graph-based neural models could be explored to enhance both accuracy and interpretability. Additionally, complementary schemes like cached registration could be leveraged to reduce signaling load and support robustness during anomalies [9].

6. Conclusions

This paper introduced a lightweight, AI-assisted anomaly detection framework tailored for cybersecurity monitoring in IMS core networks. By applying contextual statistical modeling to operational KPI data-specifically Call Setup Success Rate and Total Traffic (Erlang) - the system was able to identify both high-impact events and subtle irregularities that traditional monitoring solutions often overlook.

The results demonstrate that unsupervised anomaly detection based on z-score deviation is not only feasible but effective for uncovering signaling issues, potential security threats, and operational misconfigurations. The use of real-world data collected from multiple SBC nodes across two geographical regions validated the model's robustness and highlighted key use cases, including an anomalous traffic spike observed on March 1st.

The main advantages of this approach include the ability to detect anomalies without predefined thresholds or labeled data, compatibility with existing KPI monitoring systems and extensibility across different IMS network topologies and vendor implementations.

As 5G networks evolve and post-quantum security concerns grow in complexity, the need for adaptive, data-driven monitoring systems in telecom infrastructure becomes increasingly critical [6], [7].

References

[1]. E. E. Anderlind et al., "IMS security," in Bell Labs Technical Journal, vol. 11, no. 1, pp. 37-58, Spring 2006, doi: 10.1002/bltj.20143.

[2]. H. Pant, A. R. McGee, U. Chandrashekhar and S. H. Richman, "Optimal availability and security for IMS-based VoIP networks," in Bell Labs Technical Journal, vol. 11, no. 3, pp. 211-223, Fall 2006, doi: 10.1002/bltj.20190.

[3]. B. Soewito, O. D. Saiman and F. E. Gunawan, "Internet Protocol Multimedia Subsystem Security Risk Mitigation In Fix Telephone Network," 2019 IEEE International Conference on Engineering, Technology and Education (TALE), Yogyakarta, Indonesia, 2019, pp. 1-6, doi: 10.1109/TALE48000.2019.9225986.

[4]. A. Neureiter, Security Protection of IMS Based Telecom Networks, M.Sc. thesis, Univ. of Applied Sciences Technikum Wien, Vienna, Austria, May 2017.

[5]. A. Sardella, “Building IMS-Capable Core Networks: Backbone Foundations for Fixed-Mobile Convergence,” Juniper Networks, White Paper, Mar. 2006.

[6]. O. S. Althobaiti and M. Dohler, "Cybersecurity Challenges Associated With the Internet of Things in a Post-Quantum World," in IEEE Access, vol. 8, pp. 157356-157381, 2020, doi: 10.1109/ACCESS.2020.3019345.

[7]. H. Kim, “5G Core Network Security Issues and Attack Classification from Network Protocol Perspective,” Journal of Internet Services and Information Security (JISIS), vol. 10, no. 2, pp. 1-15, 2020.

[8]. R. Dean, W. Akpose, W. Zegeye, and F. Moazzami, “Cyber Security Architecture for Networked Telemetry,” in Proc. International Telemetering Conference (ITC), 2024.

[9]. L. Al-Doski and S. Mohan, “A Cached Registration Scheme for IP Multimedia Subsystem (IMS),” Journal of Cyber Security and Mobility, vol. 3, pp. 317-338, 2014.

Software System for Increasing Security in Telecommunications Networks

Ana-Maria NEGREI[1], Delia-Ioana LEPĂDATU[2], Gabriel PETRICĂ[1]
[1] Faculty of Electronics, Telecommunications and Information Technology,
National University of Science and Technology POLITEHNICA Bucharest, Romania
ana_maria.negrei@stud.etti.upb.ro, gabriel.petrica@upb.ro
[2] Orange Romania

Abstract

The work is based on the design of an automated software tool that provides simulation of traffic events in a telecommunications network and their analysis for the purpose of detecting Wangiri and SMS Bypass frauds. The implemented system allows parameterization of detection and prevention rules for multiple operators, for each type of fraud considered, as well as performing a complex analysis based on them for the purpose of reporting detected fraud cases. The implementation includes technologies such as Spring Boot, Java, Oracle Database, PL/SQL, React and a generative AI model, all integrated into a single architecture. The system is able to suggest measures to minimize the impact of attacks.

Index terms: telecommunications systems, SMS Bypass, SMS fraud, voice fraud, Wangiri

1. Introduction

The telecommunications industry is facing an alarming increase in fraudulent activities, with global losses estimated at $38.95 billion in 2023, a 12% increase from 2021. The rapid expansion of digital communications networks, coupled with the development of 5G technology and the use of sophisticated fraud techniques, has amplified vulnerabilities in the sector, causing significant financial losses for both operators and consumers. Telecom fraud is on the rise, with losses reaching 2.5% of the industry's total revenues. In 2024, the most common types of fraud that generate financial losses and harm to consumers include [1] (fig. 1):

- International Revenue Sharing Fraud (IRSF): estimated losses between $4 billion and $7 billion annually.
- Wangiri fraud: according to the Communications Fraud Control Association (CFCA) [2], Wangiri fraud was among the top five causes of an estimated $2.23 billion loss to the telecom industry.
- Interconnection fraud: the fraudulent manipulation of network traffic, causing annual losses of $3.11 billion.
- Account Takeover (ATO) fraud: the theft of customer identities, causing annual losses of $1.62 billion.

Recent trends indicate a rapid evolution in fraud tactics, with criminals using Artificial Intelligence (AI) and automation. AI-based scams allow fraudsters to carry out automated phishing attacks, identity fraud, and social engineering, making them difficult to detect and prevent.

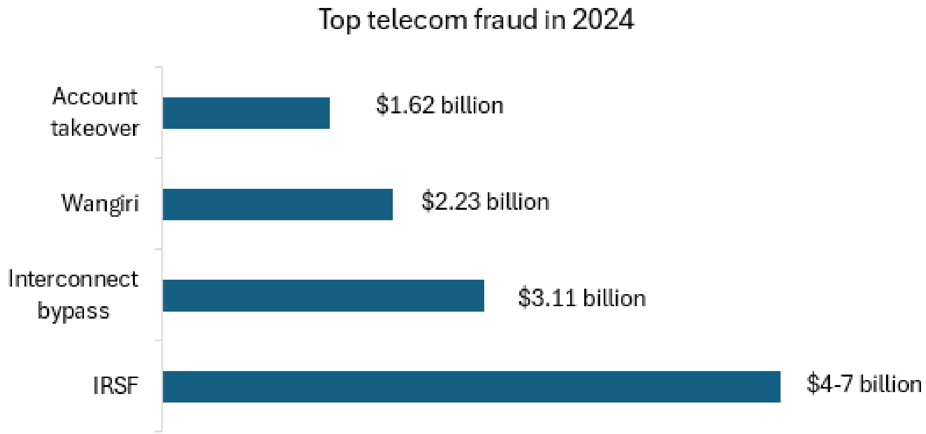


Fig. 1. Highlighting the financial impact of fraud within telecom industry

Smishing attacks in Romania increased significantly in 2023, more than six times compared to the previous year. This type of fraud mainly targets the banking sector (56%), followed by the courier industry (25%) and telecommunications services (15%) [3]. The financial impact of smishing fraud is considerable, as victims are often misled into revealing sensitive banking data or installing malware, leading to unauthorized transactions and direct financial losses. Attackers use fake websites to steal card details or gain access to users' banking applications, resulting in unauthorized access to accounts and financial theft.

In the first months of 2024, 75% of smishing cases involved identity theft by spoofing the sender's name, leading users to believe that the messages came from legitimate companies. This phenomenon affects not only individuals but also companies. Globally, 39% of consumers were the target of at least one SMS fraud attempt in 2023, highlighting the scale of this type of attack. To limit financial damage, authorities and companies emphasize the importance of user awareness and security measures, such as verifying the sender's identity, avoiding accessing suspicious links, and resetting devices if they have been compromised. Despite these efforts, smishing continues to pose a significant financial threat to individuals and organizations in Romania.

2. Fraud in telecommunications networks

Fraud Prevention Systems (FPS) are implemented to detect and mitigate fraudulent behavior using advanced technological solutions. One of the main challenges in combating fraudulent activities lies in the limitations of current infrastructure. Various mechanisms can be used before service provision, such as firewalls, encryption technologies, Subscriber Identity Modules (SIMs) in Private Branch Exchanges (PBXs), PIN codes, software-based security assessments, and customer authentication protocols. Despite their widespread use, the effectiveness of these systems has decreased over time due to the adaptive and evolving nature of attackers. In addition, fraud prevention systems are often perceived as invasive by users and are criticized for their low effectiveness. For example, assigning a security code to SIM cards is a frequently recommended measure; however, this is often neglected, leading to users forgetting the codes. Repeated and failed attempts to enter the correct code can eventually lead to the SIM card being blocked [4].

Fraud Detection Systems (FDS) are a secondary level of defense and are essential when fraudulent actions manage to bypass the implemented prevention mechanisms. These systems play an important role, as they allow for the rapid identification and notification of incidents, through a process known as real-time detection. Unlike prevention systems, which are oriented towards anticipating and blocking potential fraud, fraud detection systems operate through continuous monitoring and automated decision-making.

2.1. SMS fraud

Various types of SMS fraud have existed since the beginning of this service. Examples include SMS Bypass, SMS Malware, SMS Spam, and SMS Phishing. SMS fraud can be mitigated through multiple strategies recommended by fraud management companies, industry forums, academic researchers, and other responsible parties. Telecommunications providers are implementing these strategies, but fraud remains a problem for the industry. To effectively combat SMS fraud, it is essential to investigate the reasons why current fraud reduction methods fail.

a. SMS Spam

Represents fraud in which unsolicited messages are distributed through various electronic communication channels. These messages can be sent by individuals, companies, or even telecommunications operators. A specific form of spam, called “SMS flooding”, consists of sending messages to all users connected to an operator’s network. Such messages can lead to a breach of user privacy and data loss.

b. SMS Malware

Consists of the use of malicious software, such as viruses, worms and Trojan horses, which are intended to gain unauthorized access to data or services. This type of attack occurs when users unknowingly install infected software, which compromises their data and allows unauthorized messages to be sent to premium rate numbers. Malicious applications or scripts are designed to steal sensitive information, disrupt services, monitor user activities, access confidential data, modify device configurations or generate messages and calls to premium rate numbers without the user's consent.

c. SMS Phishing (Smishing)

It is a method comparable to phishing used in email communications on the Internet. This technique involves the attacker manipulating mobile phone users by sending forged SMS messages, using advanced social engineering practices to mislead and obtain sensitive information.

d. SMS Bypass

This type of fraud has emerged as a direct consequence of the differences in pricing between different message types and transmission routes. Traditionally, SMS messages are classified into two main categories:

- Person-to-Person (P2P): messages exchanged between individual users.
- Application-to-Person (A2P): messages sent by applications or automated systems to users, with the purpose of notifying, alerting, and sending verification codes to users.

SMS Bypass fraud is based on the pricing differences between P2P and A2P messages. While P2P messages are priced at a standard level, A2P messages - mainly used for notifications, alerts or authentication codes - are perceived as having a higher value and, implicitly, are more expensive. Taking advantage of this difference, fraudsters have developed methods to transmit A2P messages using channels intended for P2P traffic, thus bypassing official toll routes. Techniques used include:

- Using specialized equipment that simulates the sending of messages by real users.
- Accessing gray routes, i.e. unauthorized networks that avoid the control and interconnection systems of legitimate operators.
- SIM Box fraud, which involves the use of devices containing multiple SIM cards to locally retransmit international traffic.
- SMS Blasting, i.e. the transmission of many messages through systems that avoid detection and correct billing.

The consequences of SMS Bypass are multiple and affect both telecom operators and communications, as well as end users through:

- Presence of financial losses: operators suffer significant revenue losses due to fraudsters evading legitimate fees.
- Security compromise: users may receive fraudulent or malicious messages, which can lead to the theft of personal or financial information.
- Degradation of service quality: fraudulent traffic can overload networks, affecting the performance and reliability of services for legitimate users.

2.2. Voice fraud

Among the many types of fraud that exist, those that produce the greatest losses globally stand out and which persist despite attempts to combat them by operators.

a. International Revenue Sharing Fraud

IRSF represents one of the most persistent forms of fraud in the telecommunications industry, generating losses of approximately $6.7 billion in 2021, according to CFCA data [2]. Due to the complexity of the mobile network infrastructure and the large number of operators involved, this type of fraud is difficult to eliminate. It is usually orchestrated by organized groups that use illegal connections to initiate a high volume of expensive calls, thus exploiting the roaming capabilities of SIM cards. Within the IRSF, certain scenarios are inspired by the Wangiri fraud model.

b. Interconnection fraud

Interconnection fraud generates annual losses of approximately $3 billion and refers to unethical practices through which certain providers avoid paying mandatory tariffs. They manipulate the routing of calls in the network to benefit from lower costs imposed by the operator of the subscriber receiving the call, costs known in English as "termination rates".

c. Spoofing

When a call is made to another country or network, it will pass through several telephone operators. The last operator that delivers the call to the person being called charges a termination fee, similar to a commission for taking and delivering the call. This fee is paid by the calling operator to the one receiving the call.

Until 2015, termination rates for calls between European Union member states were dynamic, ranging between €0.01 and €0.15 per minute, depending on each state's pricing policies. In an effort to harmonise the European telecommunications market and reduce the costs of calls borne by users, EU regulators adopted a measure that would keep the termination rate at a fixed €0.01, similar to a rate applicable to national calls. This regulation, although beneficial for consumers, generated significant revenue losses for telecom operators, as they could no longer charge differentiated rates depending on the origin of the call.

d. Wangiri fraud

Despite their ability to innovate, fraudsters often resort to strategies that have already been tested and proven to be effective and profitable. One of these persistent methods, recognized for the advantages it offers to those who practice it, is Wangiri fraud.

The term "Wangiri" comes from the Japanese language and means "one ring and then hang up", accurately reflecting the mechanism of the scam: the caller lets the phone ring once, then hangs up. The goal of this technique, classified as a "call back" fraud, is to get users curious or concerned and call them back, usually to International Premium Rate Numbers (IPRN). These return calls generate high costs for users, while the attackers earn revenue. Typically, the calls are made during times when people are less likely to answer (at night or during business hours), to increase the chance that the call will be returned. The initial costs for the attackers are minimal, as the calls are not charged if they are not answered. Once the user calls back, the fraudsters use various methods - such as automated messages promising prizes or winnings - to keep the user on the call as long as possible. By automating the calls, they can make thousands of calls per minute. Furthermore, it is possible that there are agreements between fraudster groups and premium service providers, with the latter trying to prolong the conversations in order to increase the costs of the call and, implicitly, the joint profits [5].

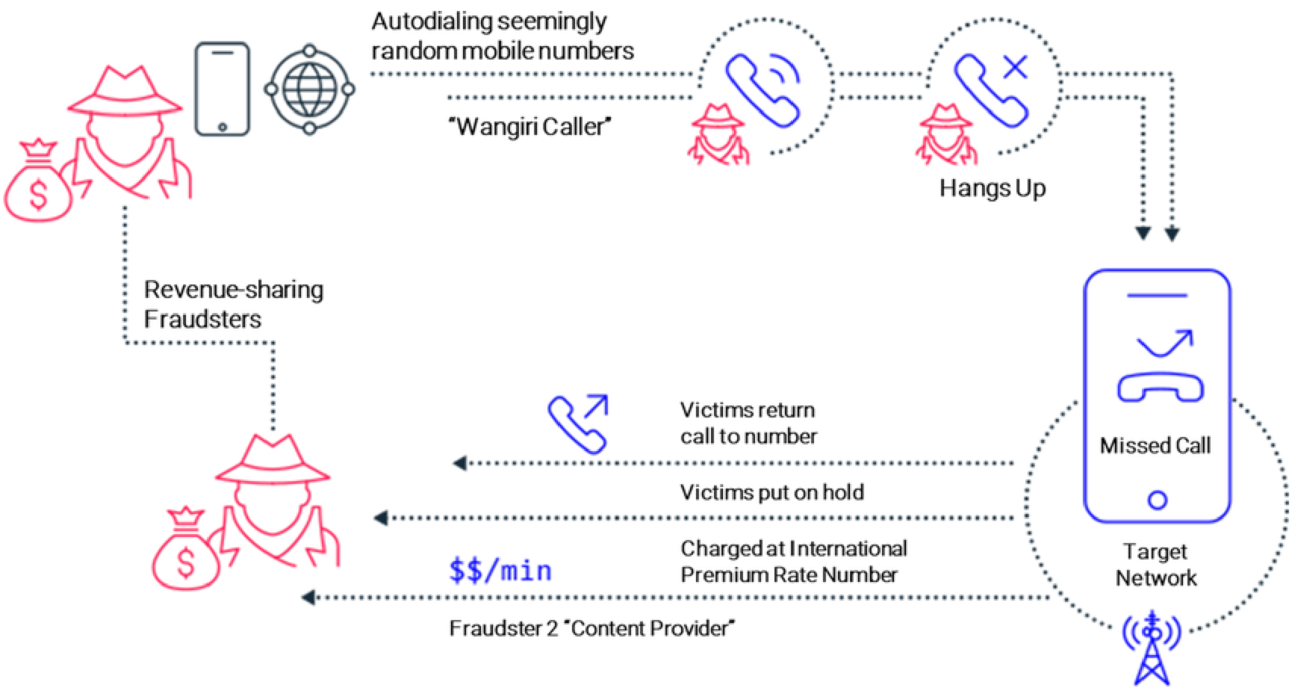


Fig. 2. Wangiri fraud method [6]

In addition to the voice call variant, there is also a form of Wangiri via SMS, in which users receive messages urging them to call back a specific number - usually the wording is urgent or alarming, in order to stimulate an immediate reaction from the victim.

The impact of this type of fraud is twofold: on the one hand, users may have to bear considerable costs, and on the other hand, operators face direct and indirect losses, such as refunding disputed amounts or damaging customer relationships. For this reason, rapid detection and blocking of suspicious calls are essential to protect both the network and users.

3. The designed detection system

In this paper, two of the most widespread and relevant types of fraud in the telecommunications sector are analyzed, selected for practical analysis: Wangiri fraud and SMS Bypass fraud. They were chosen due to their significant impact on both end users and telecommunications operators, as well as their high frequency of occurrence in modern networks.

To better understand the behavior of fraud in telecommunications networks and to evaluate the efficiency of detection mechanisms, the simulation of traffic events is an essential step in this work. Simulation allows the reproduction of fraudulent scenarios under controlled conditions, without affecting real networks and without involving authentic users, thus providing a safe environment for testing and analysis.

The main objective of the simulation is to generate relevant data that mimics real network activity, both under normal conditions (legitimate traffic) and in fraud scenarios. This data is used for:

- Analysis of fraudulent behavior, based on specific traffic patterns.
- Testing of detection algorithms, which are based on identifying anomalies in the structure and frequency of calls or messages.
- Validation of classification and filtering models, used to differentiate fraudulent from legitimate traffic.

Simulation can quickly generate large volumes of data, adapted to different types of attacks, which allows a detailed analysis of how they manifest themselves and the impact they can have on a telecommunications network.

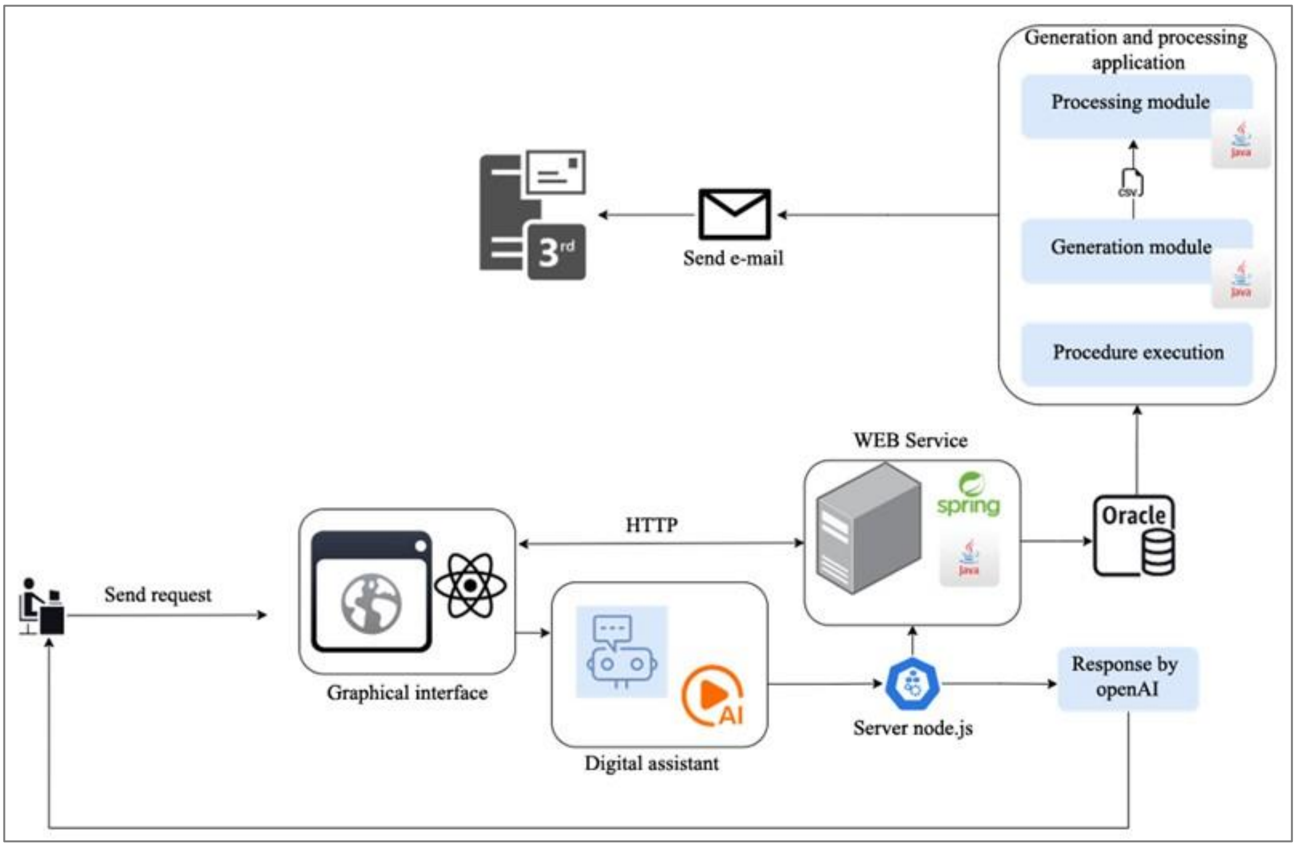


Fig. 3. Architecture of the designed system

In this work, the simulation was performed for two main types of fraud: Wangiri and SMS Bypass. The simulated events were structured to reflect real fraud scenarios as closely as possible:

- Wangiri call simulation is based on the automatic generation of a large number of short-duration events, which appear in the network as missed calls, in a short time interval to many recipients.
- SMS Bypass message simulation is based on the generation of events, imitating the fraudulent use of the infrastructure.

The simulated events are structured in the form of CDR type records, which include information such as calling number, called number, event duration, initiation time, termination time, etc. Both legitimate and fraudulent data events, calls, and messages are also simulated to provide diversity and demonstrate the functionality of data filtering procedures. This data allows for the analysis of network behavior in the presence of fraudulent activities and is the basis for testing detection solutions.

Event Generator

Creates pseudo-random events, but which follow strict rules to maintain data consistency and realism. Thus, each generated event reflects an operationally valid model, and the results can be compared with real traffic. The resulting records are saved as CSV files.

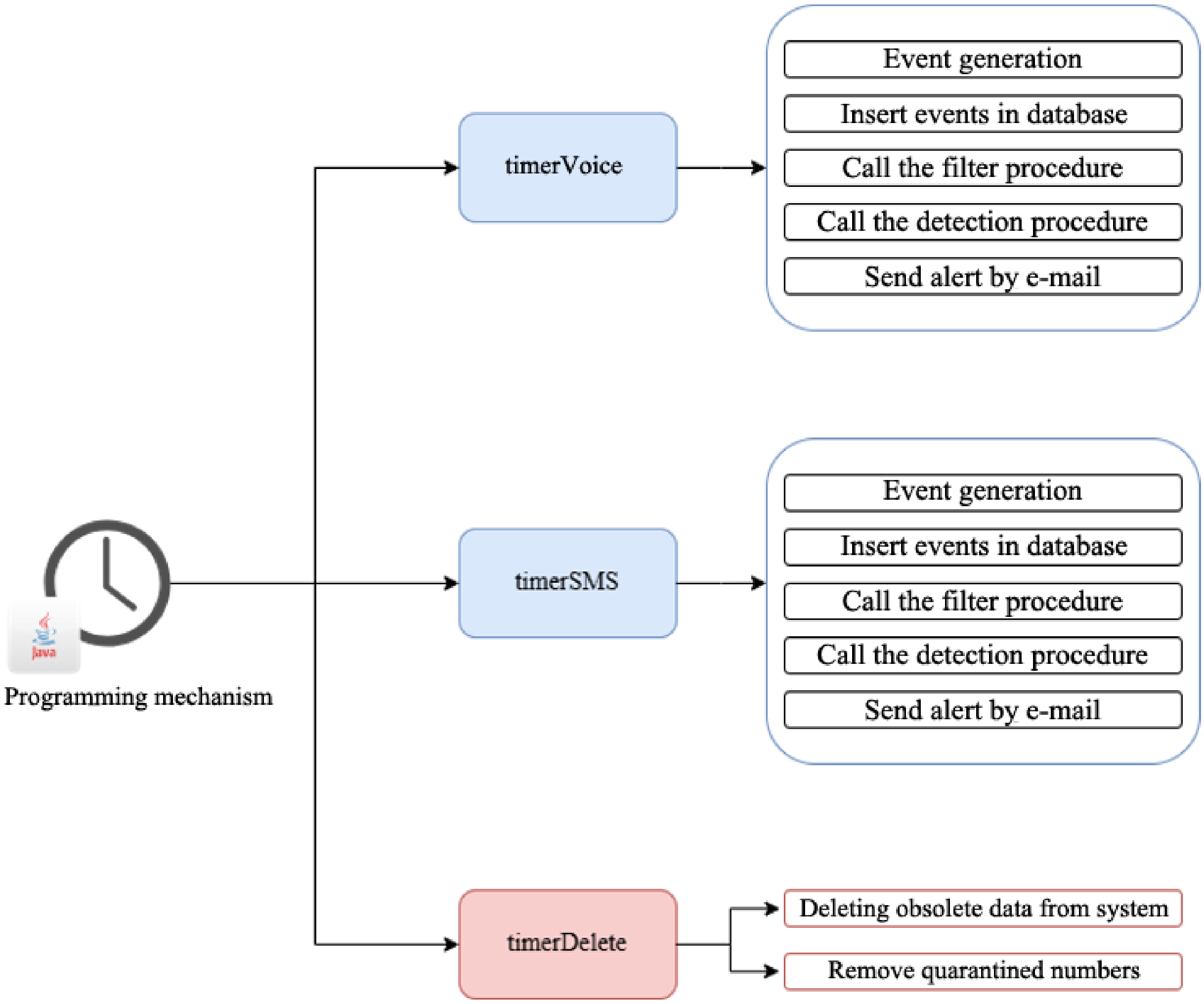


Fig. 4. Event Generator

Database

Contains tables with specific information and attributes. The relational diagram in the database is presented in Fig. 5. Next, I present in detail two of the tables implemented in the database:

detection_rules table

The "detection_rules" table defines the rules used to detect Wangiri fraud. The associated attributes are:

- id_rule - primary key, of type INTEGER, unique identifier of a detection rule.
- id_operator - reference to the operator for which the rule is applied.
- period_of_time - period (expressed in minutes) analyzed to identify fraudulent behavior.
- number_of_events - number of events recorded in the analysis period.
- number_of_distinct_destinations - number of unique destinations called.
- dispersion - degree of dispersion of calls in a certain time interval.

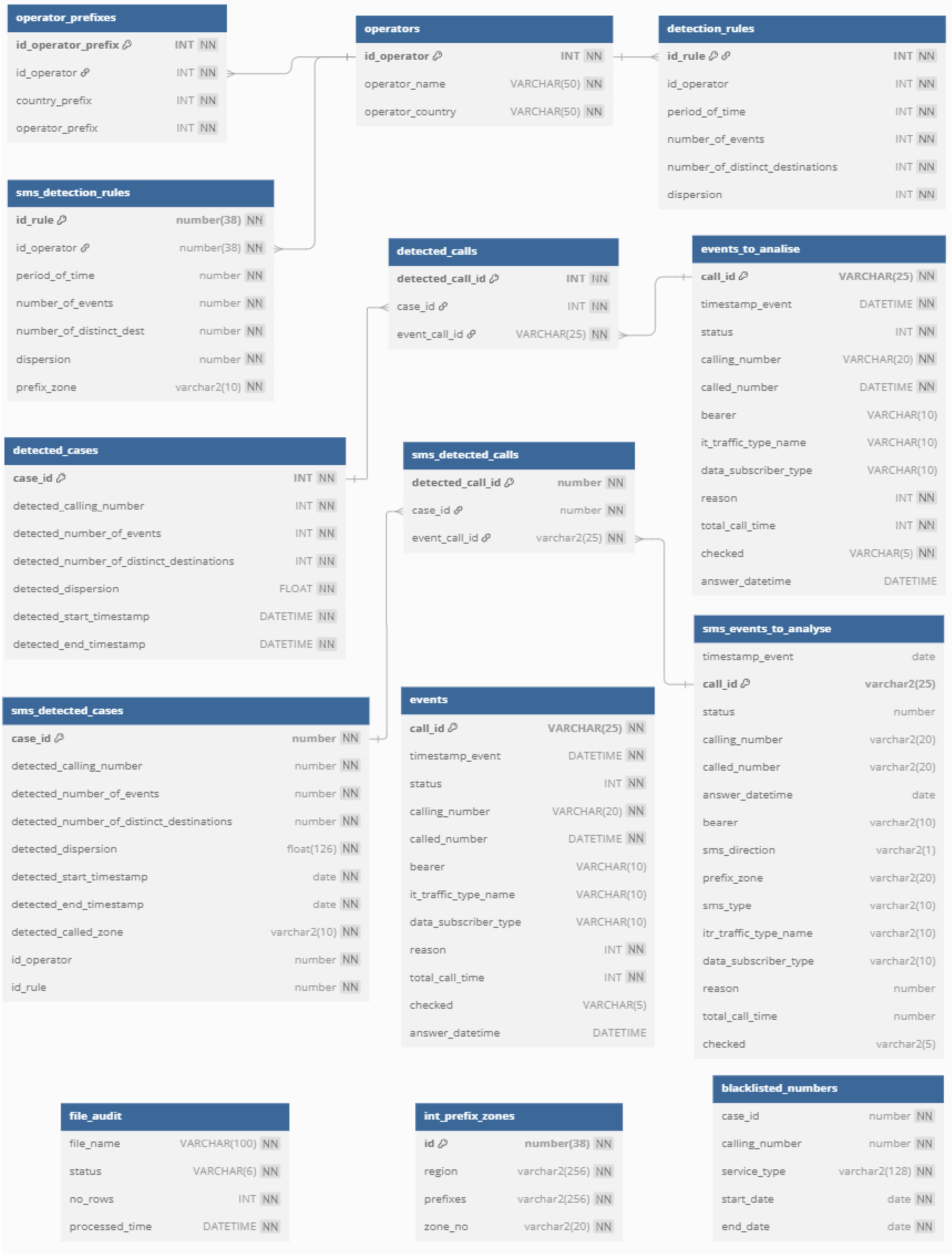


Fig. 5. The database structure

sms_detection_rules table

The "sms_detection_rules" table defines rules applied in the analysis of SMS Bypass fraud. The corresponding attributes are:

- id_rule - primary key, unique identifier of the detection rule.

- id_operator - identifier of the operator to which the rule applies.
- period_of_time - analyzed period (expressed in minutes) to identify fraudulent behavior.
- number_of_events - minimum number of SMSs that trigger the alert.
- number_of_distinct_dest - minimum number of distinct recipients.
- dispersion - dispersion of messages in a given period.
- prefix_zone - international prefix zone where messages are sent.

	ID_RULE	ID_OPERATOR	PERIOD_OF_TIME	NUMBER_OF_EVENTS	NUMBER_OF_DISTINCT_DEST	DISPERSION
1	189	285	20	6	3	0.5
2	3	2	40	10	2	0.2
3	1	1	30	10	2	0.2
4	201	285	50	2	4	0.5

(a)

	ID_RULE	ID_OPER...	PERIOD_OF_TIME	NUMBER_OF_EVENTS	NUMBER_OF_DISTINCT_DEST	DISPERSION	PREFIX_ZONE
1	1	1	10	50	6	0.12	Zone 3
2	2	1	20	30	5	0.16	Zone 2
3	3	1	30	10	3	0.3	Zone 1
4	4	1	60	5	2	0.4	Zone 0

(b)

Fig. 6. The structure of two of the tables implemented in database: (a) detection_rules and (b) sms_detection_rules

4. Conclusions

The purpose of this work was to design and implement a software system to increase the level of security in telecommunications networks. The system consists of several main components - database, Web service, graphical interface, traffic event generation and execution scheduling application and Artificial Intelligence integration logic. The system's development is based on:

- parameterization of detection and prevention rules for several operators, for each type of fraud considered.
- performing an analysis based on them for the purpose of detecting and reporting detected fraud cases.
- creating the relational model of the database, adding rules for parameterizing operators and for fraud types, defining additional tables for analysis.
- integrating a digital assistant to suggest solutions in natural language to reduce the impact of detected frauds and to provide information from the interface about operators, detected fraud cases, events associated with a fraud case, etc.
- implementation of a processing and generation application that will automate the call of detection procedures and will include two automatic event generators that will facilitate the simulation of fraud traffic for each type of attack and legitimate events.
- graphical interface for viewing and managing detected fraud cases, adding / modifying detection rules and generating statistics in CSV (Comma Separated Values) format.

The central motivation for this work was generated by the need to automate detection and prevention processes, avoiding traditional solutions based on static rules and manual processes. By combining Spring Boot, Oracle Database, React technologies and the generative artificial intelligence model, a system architecture capable of generating, filtering and analyzing traffic data, identifying suspicious patterns and providing real-time assistance to mitigate the impact of fraud was achieved.

The system evaluation demonstrated the effectiveness in identifying fraud patterns, reducing operational response time and providing visualization and interaction tools. The complete integration between data generation, automated processing and digital assistant ensures a fraud fighting platform, where each component contributes to the continuous cycle of detection, alerting and remediation.

References

[1]. Telecom Fraud Analytics: Key Trends 2024, https://dialzara.com/blog/telecom-fraud-analytics-key-trends-2024/.

[2]. CFCA 2021 Global Fraud Loss Survey, Dec 1, 2021, https://cfca.org/document/2021-fraud-loss-survey/.

[3]. R. Dumitrescu, Study: SMS scams up sixfold in Romania in 2023, https://www.romania-insider.com/study-sms-scams-sixfold-romania-2023.

[4]. Fraud Detection and Management for Telecommunication Systems using Artificial Intelligence (AI), https://0d106zzn0-y-https-ieeexplore-ieee-org.z.e-nformation.ro/stamp /stamp.jsp?tp=&arnumber=9951889.

[5]. Arafat, Mais & Qusef, Abdallah & Sammour, George. (2019). Detection of Wangiri Telecommunication Fraud Using Ensemble Learning. 330-335. DOI: 10.1109/JEEIT.2019.8717528.

[6]. Ferreira, Luís & Silva, Leopoldo & Morais, Francisco & Martins, Carlos & Pires, Pedro & Rodrigues, Helena & Cortez, Paulo & Pilastri, André. (2023). International revenue share fraud prediction on the 5G edge using federated learning. Computing. 105. 1-26. DOI: 10.1007/s00607-023-01174-w.

Trust Abuse in the Underbelly of Critical Infrastructure Operations

Eduard-Ștefan SANDU
Faculty of Applied Sciences,
National University of Science and Technology POLITEHNICA Bucharest, Romania
edy.eminem@yahoo.com

Abstract

The scientific paper presents a revolutionary cyberattack model that demonstrates how public procurement systems can be weaponized to distribute multi-extortion ransomware in critical infrastructure environments, abusing trust in legally signed documents. The attack scenario unfolds by first developing spyware capable of taking control of the digital device designed for individual use of a legitimate authorized user through which the malicious document will be signed with a qualified electronic signature, a document that will contain a ransomware. The electronically signed document will be used and sent within the framework of public procurement processes, in accordance with the rules imposed by each contracting authority through the electronic platform, named Electronic Public Procurement System. The paper is structured in sections covering the legal framework of public procurement and critical infrastructure, as well as the practical implementation scenario. The novelty of this research lies in the demonstration of a full-spectrum attack chain that combines legal compliance, identity theft and exploitation of institutional trust to bypass traditional security mechanisms.

Index terms: exfiltration, infrastructure, procurement, ransomware, signature

1. Introduction

Critical infrastructures, including but not limited to energy, water, transport, healthcare and public administration are pillars of societal stability and economic vitality. These sectors rely on a complex network of digital systems, closely linked to strict legal frameworks designed to ensure transparency, accountability and operational continuity. Central elements of critical infrastructures are the frameworks of public procurement platforms, such as the Romanian Electronic Public Procurement System, which facilitates legally binding exchanges of documents digitally signed by qualified electronic signatures issued under the eIDAS Regulation (EU) 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC and under Law no. 214/2024 on the use of electronic signatures, timestamps and the provision of trust services based on them. Although, these signatures guarantee authenticity, integrity and non-repudiation, they also introduce a new attack surface when digital identity credentials are compromised.

This scientific paper presents an innovative prototype of a cyberattack targeting critical infrastructure through document exchange within public procurement processes, weaponizing the theft of valid qualified electronic signature credentials to integrate and distribute multi-extortion ransomware into electronically signed procurement documents.

Simulating the present cyberattack requires several key code components, each tasked with executing a distinct phase of the operation. The upcoming sections offer a concise yet structured overview of these elements, emphasizing their functional roles within the broader attack sequence and setting the stage for an in-depth, phase-by-phase examination.

Initially, the attacker develops a custom spyware for an endpoint belonging to an authorized person who holds an electronic signature to sign public procurement documents. Despite multi-factor authentication and hardware protections, compromising the endpoint allows unauthorized signing with completely valid credentials.

The threat actor covertly embeds a highly sophisticated ransomware payload with a multi-extortion role in a seemingly legitimate PDF document intended for government procurement. This malicious construct features a composite architecture that includes: exhaustive cryptographic modules designed to encrypt and thereby render critical data and operational assets inaccessible, effectively inducing systemic disruption; exfiltration subroutines designed to covertly extract high-value internal information; and a multi-vector extortion paradigm. The latter transcends traditional ransom paradigms by incorporating coercive threats of public data disclosure, prolonged operational incapacitation, and reputational damage through dissemination in open markets.

Through the exploitation of illicitly procured qualified electronic signature credentials, the threat actor affixes a cryptographically binding and procedurally sanctioned digital attestation within the weaponized PDF construct, thereby imparting an ostensible veneer of juridical legitimacy consistent with prevailing regulatory frameworks. The malicious payload is subsequently transmitted via the Electronic Public Procurement System, wherein deterministic compliance mechanisms perform syntactic and structural validation of the signature's integrity absent any heuristic deviation or behavioral anomaly detection. This ostensibly legitimate dissemination channel effectively obfuscates the adversarial intent, circumventing conventional security paradigms reliant on trust anchor verification, dynamic code emulation in virtualized detonation environments and algorithmically inferred anomaly detection leveraging machine learning classifiers, thus enabling surreptitious infiltration into otherwise hardened digital infrastructures.

Recipients located within critical infrastructure entities operating under institutional trust models and subject to procedural compliance obligations, interact with the digitally signed PDF under the presumption of its authenticity and legitimacy. Upon activation, the embedded ransomware payload is covertly activated, initiating a two-stage execution process that involves cryptographically seizing data assets and covertly exfiltrating all sensitive information. This operational sequence is aligned with a multi-extortion threat paradigm, in which the adversary amplifies coercion through a confluence of data encryption, threats of unauthorized disclosure and potential reputational compromise, thereby maximizing both disruptive impact and the likelihood of ransom payment.

This scientific simulation delineates a latent yet strategically significant cybersecurity lacuna within critical infrastructure ecosystems, namely, the subversion of juridically sanctioned digital signature mechanisms as clandestine vectors for malware propagation. Diverging fundamentally from archetypal ransomware campaigns predicated on adversarial vectors such as social engineering, phishing or protocol-level exploitation, the illustrated attack modality operates entirely within the confines of regulatory compliance and procedural formality. The weaponized payload is obfuscated within a procurement document bearing a cryptographically legitimate and institutionally endorsed digital attestation, thereby engendering a high degree of epistemic trust. This sophisticated impersonation circumvents a broad spectrum of conventional defense paradigms, including certificate-based trust validation, forensic scrutiny of document provenance and heuristic signature-based endpoint protection mechanisms, ultimately facilitating undetected system ingress and operational compromise through a facade of procedural legitimacy.

Succinctly stated, the confluence of juridical infrastructures and public procurement architectures furnishes a potent vector for executing highly efficacious cyber offensives. At the core of this attack schema lies the conception and deployment of bespoke spyware, architected for covert persistence and capable of surreptitiously exfiltrating qualified electronic signature credentials, thereby subverting one of the most epistemically trusted foundations of digital identity governance. Augmenting this vector, the incorporation of a multi-faceted extortionary ransomware payload, engineered to concurrently perform cryptographic immobilization of systems and to issue credible threats of sensitive data dissemination, exponentially amplifies both psychological leverage and operational disruption. The feasibility and operational efficiency of such an attack are further enhanced through the systematic exploitation of publicly disseminated procurement intelligence, routinely published by critical infrastructure entities via electronic tendering platforms, effectively transforming regulatory transparency obligations into adversarial reconnaissance assets.

2. Public Procurement as a Weaponized Channel

Public procurement processes, carried out in accordance with the national legislative provisions of Romania, using or not the electronic platform, as the case may be, serve as essential gateways for the procurement of products, services and works to critical infrastructures. These systems increasingly rely on digital platforms and legally recognized electronic signatures to ensure the authenticity of economic operators and, where applicable, contracting authorities. However, this chapter exposes a vulnerability in how the same mechanisms that guarantee trust can be undermined by malicious actors. The chapter will focus on how attackers exploit stolen digital identities and valid electronic signatures to insert ransomware into documents transmitted through public service channels, effectively transforming public procurement workflows into weapons for critical infrastructure.

By analysing the procedural steps of public procurement through the secure e-platform, this chapter illustrates the vulnerabilities that allow the propagation of ransomware-type cyberattacks of multiple extortion under the guise of legitimate transactions of information and documents in order to participate in public procurements published by critical infrastructures. The discussion includes the technical aspects of malware integration, identity theft through spyware and manipulation of trust frameworks underlying the authenticity of public procurement. In addition, the legal and operational implications of such attacks are analysed, highlighting the gaps in existing cybersecurity and regulatory measures. This analysis reveals the urgent need for new strategies to detect, prevent and mitigate cyber threats based on public procurement that could destabilise critical sectors across Europe.

2.1. European Policies on Critical Infrastructures

The critical infrastructures of Romania, which has been a member of the European Union since 1 January 2007, have developed a comprehensive legislative framework designed to increase their resilience against cyber and hybrid threats, while ensuring the integrity of digital transactions fundamental to their functioning. This subchapter will present the most relevant EU regulations and directives and Romanian legislation.

The European Union’s regulatory framework for critical infrastructure has undergone a major transformation over the past 15 years, evolving from a limited protection model to a multi-sectoral regime focused on resilience. This legal evolution is significant given the rise in ransomware campaigns, spyware-based intrusions and hybrid cyberattacks that increasingly target critical infrastructure and exploit trust mechanisms in specific sectors of this field.

The European Union's first attempt to harmonise the security of critical infrastructures came with Directive 2008/114/EC on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection. The initial definition of critical infrastructure is found in article 2 letter (a) and (b) of Directive 2008/114/EC, as an infrastructure which is essential for the maintenance of vital societal functions, health, safety, security, social or economic well-being of people, and whose disruption or destruction would have a significant impact in a Member State as a result of the inability to maintain those functions [1].

The Directive, at the time of this scientific paper, is no longer in force, having been officially repealed by article 27 of Directive (EU) 2022/2557 of the European Parliament and of the Council of 14 December 2022 on the resilience of critical entities and repealing Council Directive 2008/114/EC, as its limited scope made it inadequate in the current threat landscape.

The Directive on the resilience of critical entities now serves as the central legislation of the European Union for the security of critical infrastructures, being in force including at the date of this scientific work, obliging member states to transpose it into national legislation.

At the same time, Directive (EU) 2022/2557 radically expands the coverage to 11 sectors, namely energy, transport, banking, financial market infrastructures, health, drinking water, wastewater, digital infrastructure, public administration, space and food production.

Directive (EU) 2022/2557, by article 2, extends the chapter on definitions, introducing a broader understanding of the field of critical infrastructures. By article 6 of the Directive, each member state shall identify critical infrastructures for the sectors and subsectors mentioned above and where critical infrastructures are identified, a member state shall take into account the results of its own risk assessment and its strategy applying several criteria cumulatively. In addition, by article 12 of the Directive, terms on risk assessment are introduced, with member states being obliged to ensure that critical infrastructures carry out a risk assessment within a period set out in the Directive and, thereafter, whenever necessary but at least once every four years, on the basis of risk assessments carried out by member states and other relevant sources of information, in order to assess all relevant risks that could disrupt the provision of their essential services. Furthermore, article 13 of the Directive mentions the obligation of critical infrastructures to adopt appropriate and proportionate resilience measures to cope with and recover from disruptions caused by physical, cyber or hybrid threats. Following the fulfilment of the previous obligations, member states shall ensure that critical infrastructures establish and implement a resilience plan or equivalent document or documents, describing the measures taken [2].

Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), is the flagship legislation of the European Union for cybersecurity risk management and significantly expands the scope, explicitly covering entities and critical infrastructures in the 11 sectors of Directive (EU) 2022/2557.

While Directive (EU) 2022/2557 focuses on the physical and operational resilience of critical infrastructures, the NIS2 Directive establishes cybersecurity obligations including critical infrastructures. Thus, article 3 of the Directive defines which structures fall under the scope of the Directive, explicitly mentioning that all identified structures will be considered universally recognized as critical infrastructures. By classifying these structures in accordance with the previous provisions, NIS2, legally obliges the identified critical infrastructures to a cybersecurity regime that recognizes their societal importance. In relation to this scientific paper, digital compromise through the use of a ransomware-type cyberattack on a critical infrastructure through public procurement, falls directly under the scope of NIS2. At the same time, article 5 of the Directive does not prevent the adoption or maintenance of provisions ensuring a higher level of cybersecurity, provided that these provisions are consistent with the obligations of the member atates under Union law. One of the main obligations of NIS2 is marked by article 7 of the Directive which requires entities, and implicitly, critical infrastructures, to implement a comprehensive set of security measures that refer to, but are not limited to, technical controls, organizational policies and supply chain security protocols. Thus, the PDF of this scientific paper injected and transmitted through a public procurement system falls into the aforementioned categories. In addition, article 21 of the Directive grants essential entities and important entities, including critical infrastructures, the power to audit, inspect and sanction, so that appropriate and proportionate technical, operational and organizational measures will be taken to manage the risks to the security of the networks and information systems that those entities use for their operations or to provide services and to prevent or minimize the impact of incidents on the beneficiaries of their services and on other services. In the event of a cybersecurity incident, critical infrastructures shall ensure that, through internal rules or other relevant legal provisions, they notify, without undue delay, the CSIRT team or, where applicable, its competent authority, of any incident that has a significant impact on the provision of their services, in accordance with the provisions of article 23 of the Directive. On the other hand, the Annexes to the Directive establish the essential and important sectors, including the 11 sectors previously defined, so that the overlap with the definitions of critical infrastructures is deliberate and complete [3].

The Digital Operational Resilience Regulation, known as the D.O.R.A., establishes a harmonised cybersecurity and risk management framework for financial entities, including a significant proportion of critical infrastructures. Under articles 5-14, the Regulation requires these critical infrastructures to adopt comprehensive domain-specific risk management frameworks that anticipate and withstand cyber disruptions. Critical infrastructures will develop internal documents covering the governance and organisation; the information and communications technology risk management framework; information and communications technology systems, protocols and tools; the identification, classification and appropriate documentation of all operational functions and all roles and responsibilities supported by information and communications technology; the protection and prevention mechanism; the rapid detection of anomalous activities; the response and recovery mechanism; backup policies and procedures and restoration and recovery procedures and methods; the method regarding lessons learned and development and communication perspectives. At the same time, for the present scientific paper on testing a ransomware attack launched through a public procurement platform, it can be carried out in accordance with the provisions of article 24 of the D.O.R.A., critical infrastructures being obliged to carry out threat-based penetration tests [4].

The Cyber Resilience Regulation, also known as the C.R.A. or EU Regulation 2024/2847, targets products and software that power modern cybersecurity operations. The regulation also requires that all digital products with direct or indirect data processing capabilities meet basic cybersecurity standards before entering the EU market. According to articles 8-12 of the C.R.A. Regulation, manufacturers must implement security by design principles, integrate vulnerability management processes, ensure timely updates and take into account critical products with digital elements, stakeholder consultation, strengthening skills in a cyber-resilient digital environment, general product safety and high-risk artificial intelligence systems [5].

Regulation (EU) 2021/696 strengthens the EU Space Programme, which sets as general priorities the provision or contribution to the provision of high-quality and up-to-date space-related data, information and services, secured where appropriate, without interruption and, wherever possible, worldwide. In addition, the Regulation also ensures the maximisation of socio-economic benefits, in particular by encouraging the development of innovative and competitive European sectors, upstream and downstream [6].

2.2. Legislative foundations of public procurement integrity

Public procurement in the European Union is not just an administrative or economic process, but a legally regulated trust mechanism designed to ensure transparency, competition and non-discrimination in the spending of public funds. It is governed by an interconnected set of European directives and national transposition laws that define every stage of the public procurement process, from the planning stage to the post-award stage.

In accordance with the provisions of article 4 of Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC, it specifies the threshold amounts, the estimated value being calculated excluding value added tax. Therefore, the contracting authority is obliged to apply the public procurement process defined in this Directive, taking into account the estimated value of the public procurement [7].

In accordance with the provisions of article 7 paragraph (7) of Law no. 98/2016 on public procurement, contracting authorities, implicitly critical infrastructures on the territory of Romania, have the right to purchase products/services/works through direct purchase offline, by publishing an advertisement within the electronic platform in the specially dedicated section, and, subsequently, the offers being received via the e-mail specified by the contracting authority [8].

On the other hand, the provisions of article 22 paragraph (2), article 62 paragraph (1) and article 137 paragraph (2) letter j) of Government Decision no. 395/2016, oblige both critical infrastructures and economic operators to sign public procurement documents with an extended electronic signature, based on a qualified certificate, issued by an accredited certification service provider, for both online and offline procurements [9].

2.3. Technical Mechanisms Underpinning Ransomware Cyberattack

The term "ransomware" is used to describe a type of malicious software that demands a ransom from a victim in exchange for the release of compromised data. The name is derived from the combination of "ransom" and "software". Payment of the ransom if often requested in cryptocurrencies like Bitcoin, but there is no guarantee that the data will be restored even after payment. Regrettably, as ransomware continues to evolve and the anonymity of the internet provides ample opportunities for exploitation, cyberattackers are able to take advantage of legal loopholes and evade detection and retribution, transforming ransomware into a highly attractive and low-risk criminal enterprise [10].

Ransomware remains one of the most profitable forms of cyberattack for financially motivated attackers, the attack being similar to a highly successful business, perfected over time and increasingly by specialized attackers, able to withstand periodic shocks and disruptions.

The cyberattack is built around implementing strong encryption for the entire targeted network and creates maximum disruption to the targeted critical infrastructure. The development of so-called multi-extortion attacks, in which attackers steal data before encryption and make it public on social media, creates an additional weak point for victims and ensures that attackers continue to have an advantage over victims who are well-prepared and can restore systems from backups. Infrastructures can be protected against these ransomware cyberattacks by using multiple layers of protection, and response tactics will be more complex. Cybersecurity is also very important, referring to measures such as ensuring that software is always up to date, having a strong password policy and using multi-factor authentication.

From the history of cyberattacks, since its creation, ransomware has evolved from simple file-locking malware to a multi-stage extortion ecosystem capable of disrupting various critical infrastructures. Moreover, modern attackers combine the attack with a combination of data theft, DDoS, service disruption and reputation damage. This change is particularly relevant for critical infrastructures and public enforcement environments, where trusted electronic communication channels are increasingly exploited to bypass defenses.

The Single-Extortion Era is the oldest era of ransomware attacks, which had low complexity, being assembled on a simple structure both in terms of design and objectives as the attack encrypts the target's files, then demands payment in cryptocurrency for the decryption keys. This specific type of ransomware attack has a technical workflow that consists of initial access vectors, payload execution, encryption process and ransom note distribution. Initial access vectors are phishing e-mails containing malicious attachments, exploiting vulnerabilities with weak and reused credentials or automated downloads from compromised websites. The encryption process consists of a hybrid system, using AES-256 for local file encryption and RSA/ECC for AES key encryption.

Since 2012, when the first ransomware-as-a-service, was developed, critical infrastructures have faced a new and powerful threat, reducing the possibility of defending against this type of attack. Key innovations of this type include modular tasks capable of disabling antivirus tools, fileless execution techniques using PowerShell and Windows Management Instrumentation, integration of command and control infrastructure to maintain persistence and signed malware binaries with stolen or forged certificates to bypass digital filters. This evolution laid the foundation for multi-layered extortion tactics, where attackers realized that simple encryption was not sufficient for coercion, especially when critical infrastructures had established backup policies and incident response frameworks.

The emergence of ransomware groups in 2019 marked a critical turning point in the history of ransomware, rewriting the possibilities and limitations of the attack. Attackers no longer relied solely on data encryption, but began stealing sensitive files and threatening to publish them on various websites, representing classified information leaks, if ransom demands were not met. The technical process of building such a cyberattack consists of initial compromise by sending phishing e-mails to deliver the spyware, data collection by using various tools for stealing credentials and exfiltrating data from the cloud, encrypting the data locally and uploading sensitive data to servers controlled by the attackers. Triple-extortion ransomware expanded beyond encryption and data leaks, introducing direct operational sabotage to amplify the pressure. One of the key techniques of such a cyberattack is the insertion of the features of DDoS attacks, representing the overwhelming of available ports, causing unavailability of services. Another key technique is represented by the vulnerability of targeting third parties by attacking the upper fora of critical infrastructure, suppliers/providers/executors and other public/private entities. In the context of quadruple extortion ransomware, attackers strategically escalate their coercive tactics beyond traditional encryption, data exfiltration and disruption of operations by introducing an additional layer, which often involves contacting third-party associates with ransom demands or other subversive tactics. Third-party associates can be but are not limited to: suppliers/providers/executors, government authorities and regulatory bodies. The latest evolution of ransomware in cyberattacks integrates automation, artificial intelligence-based recognition and cross-domain manipulation to increase influence, power or pressure on the ransom situation.

A PDF file is composed of four main sections: Header, which defines the PDF specification version; Body, containing various elements such as text, images and data streams; Xref Table, indicating the exact locations of these elements within the file and Trailer, which identifies the document's root and provides key information for interpreting its overall structure.

First, we will use pip install pikepdf, a complex file library for editing the PDF used for this scientific paper. Next, we will define the function def inject_payload(input_pdf, output_pdf): which creates a reusable function that takes the original path to the indicated file and the path to the new file after injection. To open the PDF for modification, we will use pdf = pikepdf.open(input_pdf). Next, we will create a silent JavaScript task, without alerting the victim with a message or the like. And then, we will convert the payload into a stream that encodes the JavaScript code in binary format and stores it in the PDF js_stream = Stream(pdf, malicious_js.encode("utf-8")). In addition, it is necessary to bind the task to the OpenAction command, thus, converting the JavaScript dictionary into an indirect object, we will assign it the /OpenAction properties, which means that the script will run when the PDF is opened, pdf.Root[Name("/OpenAction")]. Finally, we will save the modified PDF as a new file, ensuring that the file is prevented from becoming corrupted after saving, pdf.save(output_pdf), pdf.close().

For the purposes of this paper, the full technical implementation of the multi-extortion ransomware used in our simulation is deliberately withheld. The source code, operational tools and exploitation mechanisms constitute a large and sensitive corpus that is beyond both the scope of this paper and the limits of accepted information. Consequently, the ransomware is treated as a validated artifact used by other attackers, containing the internal mechanisms required for such malware.

2.4. Spyware as a Catalyst in Targeted Attacks on Critical Infrastructures

Among the latest trends in cyberattacks, spyware has evolved beyond passive monitoring to become a strategic enabler that facilitates complex, multi-stage intrusions, including against critical infrastructures, allowing adversaries to gain full control over targeted systems.

This paper addresses one of the most critical exploitation scenarios involving the theft of a legitimate and certified user's electronic signature credentials to electronically sign a malicious PDF on their behalf.

Furthermore, this research does not provide the full executable code of the spyware. This scenario represents a new approach to cyberattacks targeting critical infrastructures, focusing on the modeling of attacks and their use to achieve the attacker's intended goal.

Among the most important imports are subprocess to execute system commands remotely, logging to store keylogging results, threading to run tasks like keylogging and the like, requests to download external files and time to add delays between actions.

When a backdoor is created, a connection to the victim's IP address and port will be initiated. Next, an __init__ object is created that creates a stream-oriented IPv4 socket self.conn = socket.socket(socket.AF_INET, socket.SOCK_STREAM) and attempts to connect to the victim self.conn.connect((ip, port)).

The def receive(self): command is designed to retrieve data from the previously established connection. Thus, an empty string, json_result, is initialized to accumulate the received bytes in an infinite loop that attempts to read up to 1024 bytes from self.conn.recv(1024), decoding the received bytes into a string and adding them to json_result. Finally, it attempts to parse the accumulated string as an object json.loads(json_result) and returns the resulting structure after successful deserialization and if an error occurs, ValueError will allow waiting for additional received data until valid data is distributed.

The send method is responsible for sending data over the specific connections established. First, it will check if the input data is of type bytes and, if so, it will decode it into a string. Subsequently, the data is serialized into a formatted string json.dumps(data) to ensure a structured and standardized representation and finally, the string is encoded into bytes and sent over the socket self.conn.send(json_data.encode()), allowing for reliable communication with the remote host.

The execute_remote_command procedure will allow system-level commands to be executed on the host machine. It accepts a string argument command and attempts to execute it using subprocess.check_output(command, shell=True), which captures standard output as a sequence of bytes, so that, on successful execution, the captured output is returned to the caller.

In order to develop the spyware needed for this attack, we will also use change_working_dir which is designed to change the current working directory of the host process and accepts a command that is expected to contain a directory path after a command keyword. The method extracts the path component by splitting the string at the first space command.split(" ", 1)[1]). The working directory of the process is then updated via os.chdir(path), which changes the context for subsequent file system operations.

The execution of the upload_file code will allow remote data transfer to the host operating system and accepts two parameters: path, which specifies the location of the destination file and content, which represents the encoded file data. This method is used to open the target file in binary write mode with open(path, "wb"), ensuring that any existing content is overwritten. The provided content will be decoded from base64 format using the base64.b64decode(content) line of code, reconstructing the original binary data, which is then written to disk.

The use of keyscan_start is required to configure and initiate the synchronous keyboard capture routine. First, the logging subsystem will be configured via the logging.basicConfig(filename="keylog.txt", level=logging.DEBUG, format="%(asctime)s %(message)s") function, which directs time-stamped log entries to the keylog.txt file. The self.keylogging instance attribute will be set to True to indicate an active capture state.

The creds_dump function automates the recovery of stored credentials by downloading and executing a credential extraction binary from a third-party. The routine defines a URL string to obtain the remote payload via content = requests.get(url).content, so that the payload is written to disk with open("executable.exe", "wb") as a file.write(content). Finally, it will be executed by invoking result = subprocess.check_output("executable.exe all"), which runs the binary with the argument all and captures the standard output in the variable result.

The get_ip_address function will determine the outgoing IPv4 address of the host machine by creating a socket and querying the local endpoint of the socket. Specifically, a datagram socket is instantiated with s = socket.socket(socket.AF_INET, socket.SOCK_DGRAM), where socket.AF_INET specifies IPv4 and socket.SOCK_DGRAM specifies a connectionless UDP socket.

Finally, the last sequence determines the outgoing IPv4 address of the host by calling the function ip = get_ip_address(). The Backdoor class will, then, be instantiated with that address and the literal port 4444 via the function backdoor = Backdoor(ip, 4444), causing the constructor to create an IPv4 TCP socket and attempt a connection to (ip, port). At the same time, backdoor.run() is invoked to start the instance's runtime behavior.

3. Analyzing threat vectors through empirical study of cyberattack simulation

All data presented in this chapter is for informational purposes only, including screenshots and referenced materials, are extracted exclusively from public sources, not protected by specific laws and do not reveal classified information or content restricted by law. Moreover, the data presented in this chapter, including screenshots, are provided in the national language of Romania, namely Romanian. The simulation and testing of the cyberattack were carried out exclusively on the national public procurement platform, using only information accessible to the Romanian public.

A controlled ethically constrained simulation of a cyberattack using the public procurement electronic platform will be presented. Adopting an attacker-centric analytical stance, we conceptually reconstruct the attacker's decision-making process to expose how trust mechanisms within the public procurement sector can be abused and where defensive controls fail. For clarity and safety, the offensive capabilities are treated as a model that does not provide full exploit code, operational payloads or the like.

The scenario is developed in distinct phases, useful from an attacker's analytical point of view, described only at a conceptual level sufficient to obtain observable indicators and mitigation hypotheses. The phases include, but are not limited to: targeted recognition and victim profiling to identify a valid and certified electronic signature; design of social-engineering artifacts that model the informational profile necessary to construct a persuasive phishing lure and the subsequent abuse of a compromised qualified electronic signature to confer ostensible legitimacy on a weaponized document; selection of a contracting authority on the SEAP platform that has the character of critical infrastructure; controlled sending of an instrumented document that represents an offer for application to a public procurement process whose activation triggers the ransomware.

The simulation can be executed by any natural or legal person, in isolated and instrumented environments, under institutional supervision for the purpose of substantiating the defensive strategy and incident response, without training malicious activities.

The initial phase of the modeled scenario corresponds to an open-source reconnaissance carried out using equipment dedicated to public procurement. Within the Electronic Catalogue of the SEAP platform, a prospective target profile can be built from the information that economic operators choose to publish, including the e-mail address that will be used to transmit the spyware.

Informatii ofertant

Ofertant: S.C. [redacted] S.R.L. CIF: [redacted]

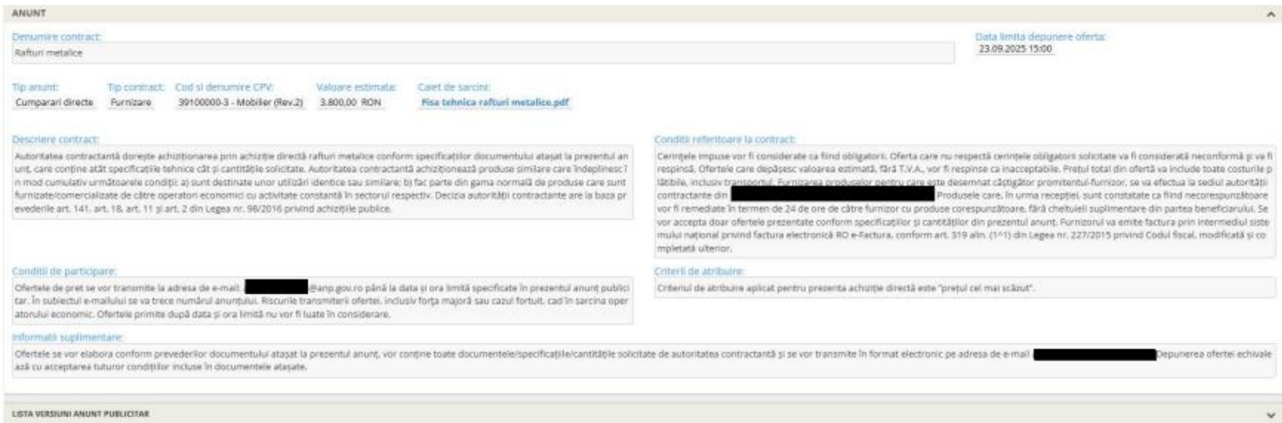
Adresa: [redacted] Teleorman, Localitate: Alexandria, Cod postal: [redacted] Tara: Romania

Website: - Tel: [redacted] Fax: [redacted] E-mail: [redacted]@yahoo.ro

Fig. 1. Possible victim information within the SEAP platform

Next, in the second phase, the attacker focuses on the design of social engineering content aimed at causing the victim to reveal the signing credentials. The scenario proposed for this scientific paper is an offer to enroll in validated specialized courses in public procurement for the position of expert. After accessing the transmitted PDF, the spyware will connect to the victim's device, according to the previous mentions regarding the spyware's attributions.

All simulation artifacts referred to in this chapter are synthetic or come from public, non-confidential materials in Romanian from the national procurement platform (SEAP) and do not provide operational guidance for exploitation. As a result, the model is oriented towards cybersecurity against an attacker attempting to use a public procurement workflow as an entry vector. The simulation models a scenario in which a public procurement advertisement is published by an entity classified as a critical infrastructure operator. The advertisement indicates that the bid must be submitted via electronic means, namely e-mail, and requires that it include technical specifications and total bid value. The subject of the e-mail will specify the advertisement number to increase the success rate of the attack, such as ADV14991XX in Figure 2.



ANUNT

Denumire contract: Rafturi metalice

Data limita depunere oferta: 23.09.2025 15:00

Tip anunt: Cumparari directe | Tip contract: Furnizare | Cod si denumire CPV: 39100000-3 - Mobilier (Rev.2) | Valoare estimata: 3.800,00 RON | Caiet de sarcini: Fisa tehnica rafturi metalice.pdf

Descriere contract:
Autoritatea contractantă doreşte achiziţionarea prin achiziţie directă rafturi metalice conform specificaţiilor documentului ataşat la prezentul anunt, care conţine atât specificaţiile tehnice cât şi cantităţile solicitate. Autoritatea contractantă achiziţionează produse similare care îndeplinesc în mod cumulativ următoarele condiţii: a) sunt destinate unor utilizări identice sau similare; b) fac parte din gama normală de produse care sunt furnizate/comercializate de către operatori economici cu activitate constantă în sectorul respectiv. Decizia autorităţii contractante are la baza prevederile art. 141, art. 18, art. 11 şi art. 2 din Legea nr. 98/2016 privind achiziţiile publice.

Conditii referitoare la contract:
Cerinţele impuse vor fi considerate ca fiind obligatorii. Oferta care nu respectă cerinţele obligatorii solicitate va fi considerată neconformă şi va fi respinsă. Ofertele care depăşesc valoarea estimată, fără T.V.A., vor fi respinse ca inacceptabile. Preţul total din ofertă va include toate costurile plătibile, inclusiv transportul. Furnizarea produselor pentru care este desemnat câştigător promitentul-furnizor, se va efectua la sediul autorităţii contractante din [redacted] Produsele care, în urma recepţiei, sunt constatate ca fiind necorespunzătoare vor fi remediate în termen de 24 de ore de către furnizor cu produse corespunzătoare, fără cheltuieli suplimentare din partea beneficiarului. Se vor accepta doar ofertele prezentate conform specificaţiilor şi cantităţilor din prezentul anunţ. Furnizorul va emite factura prin intermediul sistemului naţional privind factura electronică RO e-Factura, conform art. 319 alin. (1^1) din Legea nr. 227/2015 privind Codul fiscal, modificată şi completată ulterior.

Conditii de participare:
Ofertele de pret se vor transmite la adresa de e-mail: [redacted]@anp.gov.ro până la data şi ora limită specificate în prezentul anunţ publicitar. În subiectul e-mailului se va trece numărul anunţului. Riscurile transmiterii ofertei, inclusiv forţa majoră sau cazul fortuit, cad în sarcina operatorului economic. Ofertele primite după data şi ora limită nu vor fi luate în considerare.

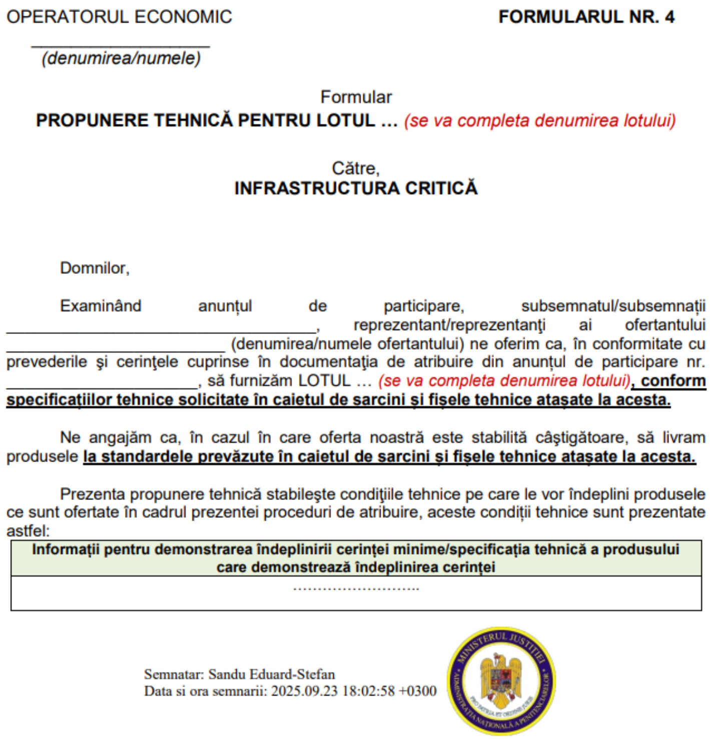
Criterii de atribuire:
Criteriul de atribuire aplicat pentru prezenta achiziţie directă este "preţul cel mai scăzut".

Informatii suplimentare:
Ofertele se vor elabora conform prevederilor documentului ataşat la prezentul anunţ, vor conţine toate documentele/specificaţiile/cantităţile solicitate de autoritatea contractantă şi se vor transmite în format electronic pe adresa de e-mail [redacted] Depunerea ofertei echivalează cu acceptarea tuturor condiţiilor incluse în documentele ataşate.

LISTA VERSIUNI ANUNT PUBLICITAR

Fig. 2. Critical infrastructure advertisement within the SEAP platform

In the final step, the study models the effect of a weaponized bidding document, electronically signed using previously compromised credentials. For ethical and legal reasons, the paper does not describe how the developed ransomware is integrated or activated, rather, it treats the malicious functionality as an event that can be triggered when a signed bid is opened in a critical infrastructure environment. In case critical infrastructure personnel do not trust the submitted document, noticing the extension and properties of a valid electronic signature attached to the document will increase their confidence in opening the document, as illustrated in Figure 3.



OPERATORUL ECONOMIC **FORMULARUL NR. 4**

____
(denumirea/numele)

Formular
PROPUNERE TEHNICĂ PENTRU LOTUL … *(se va completa denumirea lotului)*

Către,
INFRASTRUCTURA CRITICĂ

Domnilor,

Examinând anunţul de participare, subsemnatul/subsemnaţii ______________________________, reprezentant/reprezentanţi ai ofertantului _________________________ (denumirea/numele ofertantului) ne oferim ca, în conformitate cu prevederile şi cerinţele cuprinse în documentaţia de atribuire din anunţul de participare nr. ______________________, să furnizăm LOTUL … *(se va completa denumirea lotului)*, **conform specificaţiilor tehnice solicitate în caietul de sarcini şi fişele tehnice ataşate la acesta.**

Ne angajăm ca, în cazul în care oferta noastră este stabilită câştigătoare, să livram produsele **la standardele prevăzute în caietul de sarcini şi fişele tehnice ataşate la acesta.**

Prezenta propunere tehnică stabileşte condiţiile tehnice pe care le vor îndeplini produsele ce sunt ofertate în cadrul prezentei proceduri de atribuire, aceste condiţii tehnice sunt prezentate astfel:

Informaţii pentru demonstrarea îndeplinirii cerinţei minime/specificaţia tehnică a produsului care demonstrează îndeplinirea cerinţei
……………………..

Semnatar: Sandu Eduard-Stefan
Data si ora semnarii: 2025.09.23 18:02:58 +0300

Fig. 3. Electronically signed offer template

4. Conclusions

This study examined a simulated cyberattack chain in a secure and ethically constrained environment, where credential compromise enabled by spyware and abuse of qualified electronic signatures are used to weaponize workflows against entities classified as critical infrastructure. Based on the simulation, we abstracted the offensive capability and focused exclusively on observable indicators, attack architectures and legal implications.

Trust-based channels in public procurement, especially digitally signed documents accepted without contextual validation, create opportunities for adversaries. When combined with credential harvesting via spyware, these channels significantly increase the likelihood of an attacker achieving significant compromises.

The ability to transform exposure to reputational and regulatory risks into a multi-extortion cyber weapon substantially increases both the tactical and strategic damage of successful intrusions.

Conventional signature verification provides cryptographic assurance of integrity and origin, but does not attest to the signer's intent, the provenance of the device or the appropriate contextual character of a signing event, conditions that adversaries can exploit.

Measures such as enhanced authentication for signing events, hardware-bound keys and strong alignment of legal provisions with internal rule implementation, demonstrably, reduce simulated success rates, but they also introduce operational difficulties and potential false positives that need to be managed.

Operators of national public procurement systems should treat information input into the procurement process flow and the processing of signed documents as critical security control points requiring increased protection and monitoring.

References

[1]. Official Journal of the European Union “Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the

assessment of the need to improve their protection" eur-lex.europa.eu. Accessed: July 29, 2025. [Online.] Available: https://eur-lex.europa.eu/eli/dir/2008/114/oj/eng.

[2]. Official Journal of the European Union "Directive (EU) 2022/2557 of the European Parliament and of the Council of 14 December 2022 on the resilience of critical entities and repealing Council Directive 2008/114/EC" eur-lex.europa.eu. Accessed: July 29, 2025. [Online.] Available: https://eur-lex.europa.eu/eli/dir/2022/2557/oj/eng.

[3]. Official Journal of the European Union "Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive)" eur-lex.europa.eu. Accessed: July 29, 2025. [Online.] Available: https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng.

[4]. Official Journal of the European Union "Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011" eur-lex.europa.eu. Accessed: July 29, 2025. [Online.] Available: https://eur-lex.europa.eu/eli/reg/2022/2554/oj/eng.

[5]. Official Journal of the European Union "Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act)" eur-lex.europa.eu. Accessed: July 29, 2025. [Online.] Available: https://eur-lex.europa.eu/eli/reg/2024/2847/oj/eng.

[6]. Official Journal of the European Union "Regulation (EU) 2021/696 of the European Parliament and of the Council of 28 April 2021 establishing the Union Space Programme and the European Union Agency for the Space Programme and repealing Regulations (EU) No 912/2010, (EU) No 1285/2013 and (EU) No 377/2014 and Decision No 541/2014/EU" eur-lex.europa.eu. Accessed: July 29, 2025. [Online.] Available: https://eur-lex.europa.eu/eli/reg/2021/696/oj/eng.

[7]. Official Journal of the European Union "Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC" eur-lex.europa.eu. Accessed: July 29, 2025. [Online.] Available: https://eur-lex.europa.eu/eli/dir/2014/24/oj/eng.

[8]. Official Gazette "Law No. 98 of 2016 on public procurement" cdep.ro. Accessed: July 29, 2025. [Online.] Available: https://www.cdep.ro/pls/legis/legis_pck.htp_act?ida=137225.

[9]. Official Gazette "Government Decision No. 395/2016 for the approval of the Methodological Norms for the application of the provisions relating to the award of the public procurement contract/framework agreement in Law No. 98/2016 on public procurement." cdep.ro. Accessed: July 29, 2025. [Online.] Available: https://www.cdep.ro/pls/legis/legis_pck.lista_mof?idp=25750.

[10]. Eduard-Ștefan SANDU, "Prevention of Widespread Ransomware Cyber-Attacks through the SEAP Platform" in Proceedings of the International Conference on Cybersecurity and Cybercrime (IC3), Volume X, Romania: RAISA, 2023, pp. 230-240.

Enhancing the Security of High-Responsibility Information Systems Through Fault Tree Modeling

Constantin-Alin COPACI, Ioan C. BACIVAROV
Faculty of Electronics, Telecommunications and Information Technology,
National University of Science and Technology POLITEHNICA Bucharest, Romania
constantin.copaci@stud.etti.upb.ro, ioan.bacivarov@upb.ro

Abstract

With the advancement of technology, high-dependability information systems have become indispensable for the efficient operation of activities in strategic sectors, among which communications play a crucial role. This paper aims to ensure the security and reliability of a communication system through the application of fault tree analysis. This approach seeks to identify vulnerable components, assess the probability of adverse events within the system, and propose measures to enhance its performance under conditions of risk or failure.

Index terms: communication system, failure, Fault Tree, reliability, security

1. Introduction

High-dependability information systems (HDIS) are technical or computer-based systems whose correct, continuous, and secure operation is essential for preventing serious consequences that could affect human life, the environment, public safety, or critical infrastructure.

Ensuring security, reliability, and fault tolerance constitutes a fundamental requirement for high-dependability information systems, as any malfunction or incident may have significant repercussions. In this context, Fault Tree Analysis (FTA) serves as a key tool, as it enables the logical and structured modeling and assessment of potential causes leading to system failures [1], [2].

2. Fault Tree modeling for high-dependability information systems

Fault Tree modeling represents a fundamental stage in the reliability and security analysis of high-dependability information systems, as it enables the identification and assessment of factors that may lead to failures or undesired events. This section presents the main stages involved in constructing a fault tree, as well as specific adaptations relevant to information systems [3], [4].

The Fault Tree Analysis (FTA) method is a systematic analytical technique used to identify and evaluate the potential causes of undesired events or failures within complex systems. It allows for the logical modeling of cause-effect relationships and is widely applied in the fields of safety and reliability, particularly for high-impact critical systems. The primary objective of this method is to identify all possible combinations of faults or events that may lead to the main undesired outcome, referred to as the top event.

Thus, risks can be assessed, and preventive or corrective measures can be prioritized. The main operating principles are as follows:

- The method begins with a specified undesired event (for example, a system failure or a security breach).

- A reverse logical tree is then constructed, where the top event represents the root of the tree.
- The events and conditions that may lead to this top event are represented as nodes and are connected using logical operators (AND, OR, etc.).
- The basic events (fundamental elements that cannot be further decomposed into other causes) are identified.
- The analysis can be qualitative (determining the failure paths) or quantitative (calculating probabilities and frequencies).
- The construction of a fault tree involves a series of systematic steps designed to ensure a complete and accurate modeling of the causes leading to the top event:
- Definition of the top event: clearly establishing the malfunction or undesired incident to be analyzed.
- Identification of immediate causes: determining the events that can directly lead to the top event.
- Decomposition into intermediate and basic events: continuing the analysis until the fundamental events, which cannot be further broken down, are identified.
- Application of logical operators: connecting the events using logical operators (AND, OR, etc.) to describe the causal relationships.
- Validation of the fault tree: verifying the consistency and completeness of the model through expert consultation and critical review.

Within information systems, both fundamental and intermediate elements can be diverse and complex. A hardware malfunction [4], [5] may appear simple, yet when combined with a software error or human mistake, the resulting effects can be devastating. Each component-from servers to networks, from software code to operator procedures-must be regarded as a crucial element within this fragile network.

To provide an accurate and comprehensive representation, modeling does not stop at identifying causes but also incorporates their probabilistic assessment. This adds dimensions related to how vulnerable a system is to attacks, how frequently hardware or software errors may occur, and how effective the response to incidents proves to be.

This complex integration of factors enables not only an understanding of how and why failures occur, but also the anticipation and prevention of such events, thereby transforming the analysis into a vital tool for safeguarding and maintaining the functional integrity of highly critical systems.

3. Case Study - Fault Tree Analysis applied to a communication network within an organization

In communication networks, Fault Trees are employed to analyze the network's availability, reliability, and security. In such networks, cyberattacks can be modeled as basic events, representing various methods through which security may be compromised [5], [6], [7].

Examples:

- Unauthorized access: The attacker gains access to network resources through weak passwords, phishing, or other social engineering methods.
- DDoS attack (Distributed Denial of Service): Overloading network resources with false traffic to block legitimate user access.
- Software vulnerability exploitation: The attacker leverages known vulnerabilities (such as unpatched software) to execute malicious code.
- Data interception (Man-in-the-Middle): Capturing and manipulating network traffic.
- Malware and ransomware: Installing malicious programs that disrupt operations or compromise data.

Each type of attack can be considered an individual basic event in the Fault Tree.

Technical faults represent hardware or software issues that can affect the proper functioning of the network and may lead to service interruptions or performance degradation. Examples of such faults include:

- Hardware failure: Malfunction of routers, switches, servers, or other physical equipment.
- Software error: Bugs or faults in operating systems, firmware, or network applications.
- Connection loss: Physical disconnection or interruptions in cables, optical fibers, or wireless signals.
- Network overload: Excessive traffic leading to increased latency or packet loss.
- Configuration errors: Incorrect settings in devices or security policies.

These technical faults also represent basic events and can be combined with cyberattacks within the fault tree to evaluate the overall risk of the communication network.

Within an organization, the communication infrastructure is essential for the execution of daily operations. The network consists of a mix of hardware devices (routers, switches, servers) and software components (operating systems, applications, security policies), all interconnected and exposed to both technical and cybersecurity risks

To assess the overall risk of service disruption (considered the top event), a Fault Tree Analysis (FTA) approach was employed. Each type of attack or technical fault is treated as a basic event that may contribute to system failure.

The objective of this case study is to identify and model the primary potential causes that could lead to a complete IT service outage within the organization, with the goal of establishing priorities for preventive and corrective measures.

Accordingly, using Matlab software [8], a cyberattack is modeled as a Fault Tree. The probability of compromise is calculated both analytically (using formulas) and through Monte Carlo simulation, and the impact of changes in individual probabilities on the overall risk is analyzed. Additionally, the fault tree is visually represented. [9], [10],[11]

We consider the compromise event E0 =E1 OR (E2 AND E3), where E1 represents a successful phishing attack; E2 corresponds to the exploitation of a software vulnerability; represents unauthorized access (physical or remote) (Figure 1).

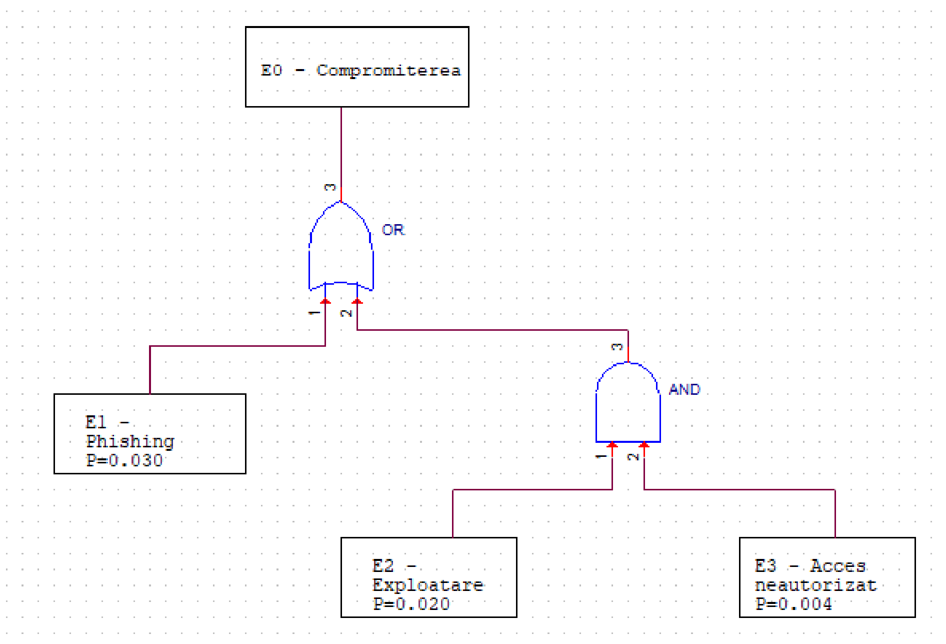


Fig. 1. Compromise Event

The basic probabilities for each elementary event are defined (Figure 2).

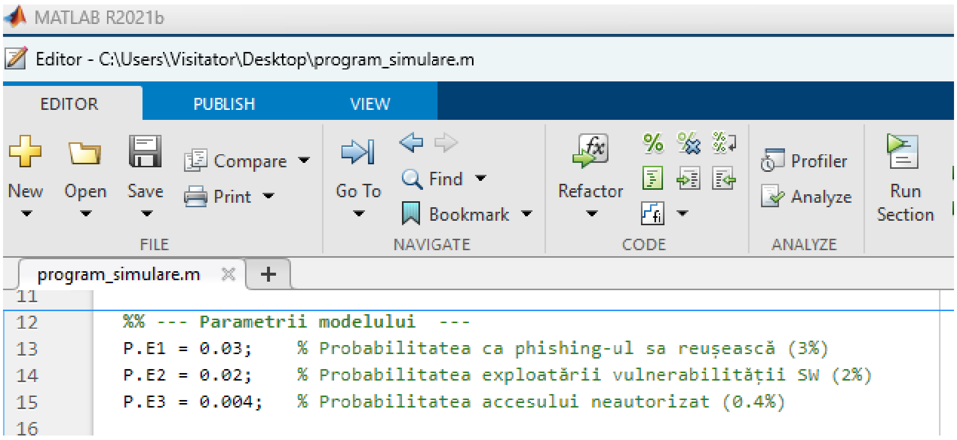


Fig. 2. Model Parameter Definition

The system is compromised if phishing occurs (E1) or if the software vulnerability is exploited while there is also unauthorized access (E2 AND E3).

Assuming the three events are independent, the combined probability for E0 (compromise) is calculated analytically as follows:

$$P(E2\ AND\ E3) = P(E2) \times P(E3) = 0.02\ \times 0.004 = 0.00008$$

$$P(E0) = P(E1)\ OR\ P(E2\ AND\ E3)$$

$$P(E0) = 1 - \big(1 - P(E1)\big) \times \big(1 - P(E2) \times P(E3)\big) = 1 - (1 - 0.03) \times (1 - 0.00008) \approx 0.3008$$

The probability of compromise is approximately 3.008%, which is nearly equal to the probability of phishing, since P(E2 AND E3) is much smaller (0.008%).

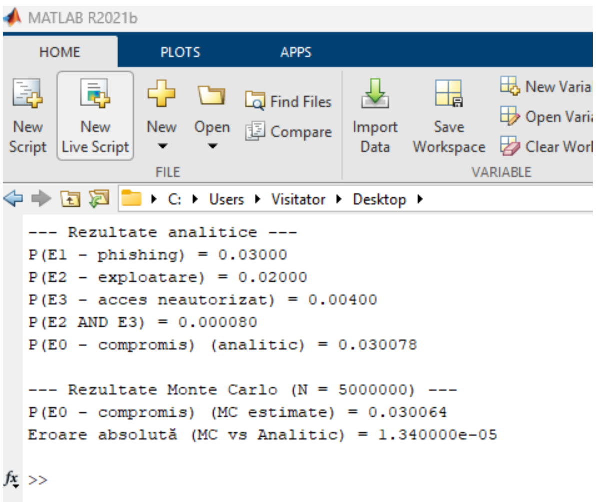


Fig. 3. Analytical Results and Monte Carlo Results

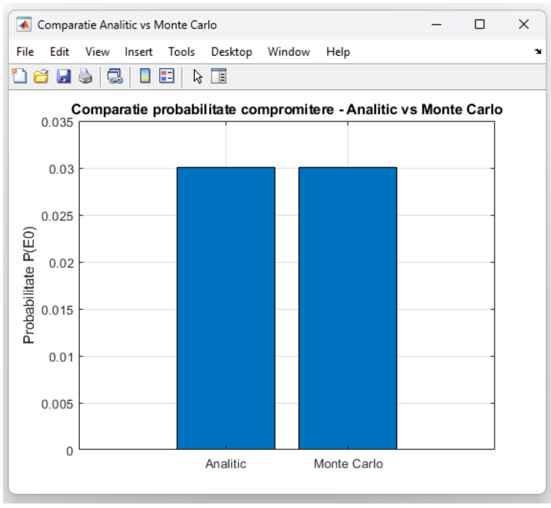


Fig. 4. Compromise Probability Comparisons - Analytical vs. Monte Carlo

Since the error between the Monte Carlo simulation and the analytical calculation is very small, it follows that the simulation correctly validates the model. The Monte Carlo estimate can be considered a verification of the analytical formula.

We further analyze how the probability of compromise, P(E0), varies when:

- P(E1) changes from 0 to 10%.
- P(E2) changes from 0 to 10%, keeping P(E3) constant.

If the curve for P(E1) is steeper, the system is more sensitive to phishing. If the curve for P(E2) is steeper, the software vulnerability has a greater impact. This allows us to identify which component of the fault tree is critical for the system's security (Figure 5).

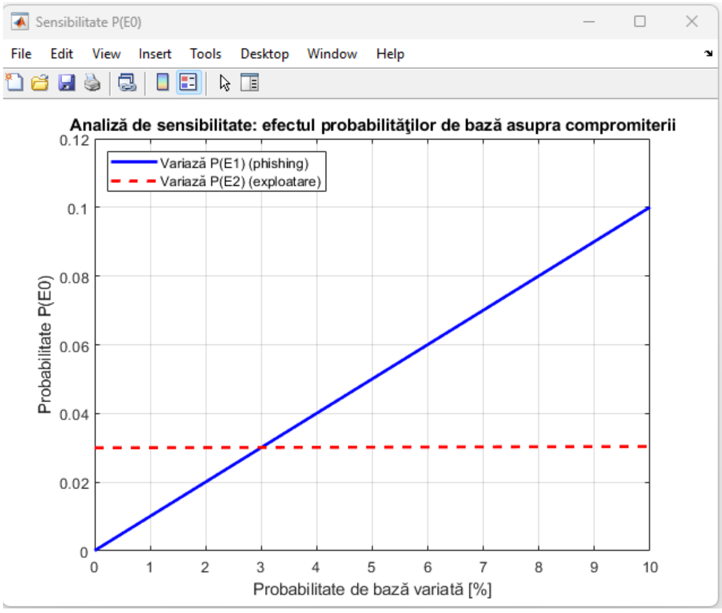


Fig. 5. Sensitivity Analysis

A simplified directed graph is constructed for the fault tree with the following nodes: E0 (compromise); OR (logic gate); E1 (phishing); AND (logic gate); E2 (exploitation); E3 (unauthorized access).

Figure 6 shows the fault tree graph, where the probabilities are also displayed next to the basic events.

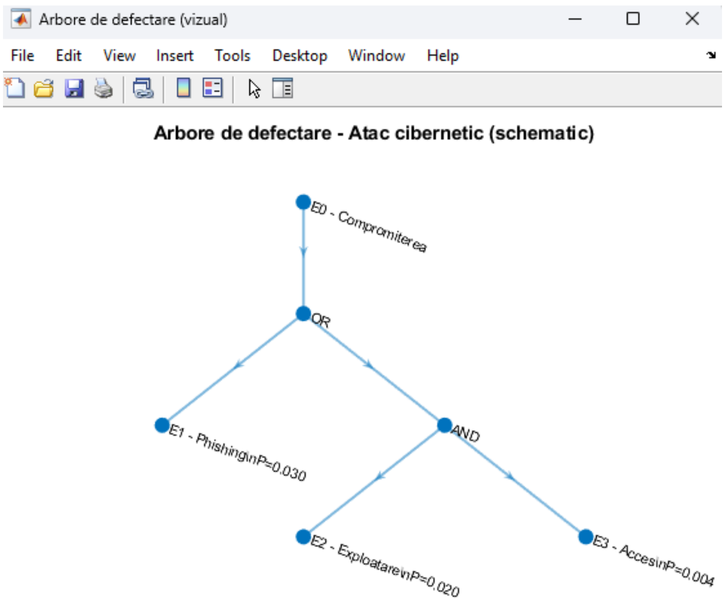


Fig. 6. Fault Tree - Cyberattack

From the conducted analyses, it is observed that the probability of compromise is dominated by phishing (E1), since P(E1) = 3% is much higher than the probability of E2 AND E3 = 0.008%.

Thus, the combined probability of exploiting the software vulnerability and unauthorized access is low, but still relevant enough to be considered in the overall system risk assessment.

The model provides an appropriate analytical framework for performing a sensitivity analysis, investigating the impact of variations in the individual probabilities of basic events (phishing and software exploitation) on the system's compromise probability.

The results obtained through Monte Carlo simulation offer a robust numerical validation of the analytical calculation, confirming the accuracy of the estimated probability for the top event based on the independence assumption.

The schematic graphical representation of the fault tree facilitates understanding of the logical relationships between basic events and how they combine to generate the top event, providing an intuitive view of the risk structure.

The following table presents, in a structured way, some protective measures that can be implemented according to each event [12], [13]:

Table 1. Protective Measures for Each Event

Event Code	Event Description	Probability	Protective Measures
E1	Phishing	0.03	- Anti-phishing training - Phishing simulations - Email filtering- MFA
E2	Exploitation of vulnerabilities	0.02	- Patch management - Vulnerability scanning - Intrusion prevention
E3	Unauthorized access	0.004	- Multi-Factor Authentication - Role-based access control - Logging & monitoring - Network segmentation
E0	System compromise	≈ 0.03008	- Integrated cybersecurity - Response plan - Periodic audits

4. Conclusions

This study has demonstrated the importance and effectiveness of the Fault Tree Analysis (FTA) method in assessing the security and reliability of high-functionality information systems (HFIS) used in critical sectors with significant impact on essential infrastructures and public safety.

By modeling cause-and-effect relationships logically, the fault tree allows for the clear identification of basic events and their combinations that may lead to the compromise of a critical system, providing a rigorous framework for risk assessment. The case study applied to a communication system highlighted how cyberattacks and technical failures can be integrated into a coherent probabilistic model.

Analytical results, obtained under the assumption of independent elementary events, were validated through Monte Carlo simulation, which confirmed the system compromise probability estimate with minimal error. This correlation strengthens confidence in the applicability of the method for quantitative security analysis.

Sensitivity analysis revealed that the risk of system compromise is dominated by the probability of phishing, while the combined probability of software vulnerability exploitation and unauthorized access, although less likely, remains an important factor that should not be overlooked in the overall risk assessment.

The schematic graphical representation of the fault tree facilitated understanding of the interdependencies among events and the logical structure of risks, serving as a valuable tool for communication and decision-making within security and reliability management processes.

Based on these findings, it is recommended to incorporate fault tree methodology into risk assessment and management strategies for critical information systems, as well as to extend the model by including dependencies between events and using advanced simulation techniques to improve estimation fidelity.

In conclusion, fault tree modeling proves to be a robust, adaptable, and relevant methodological tool for ensuring the security and reliability of high-functionality information systems, substantially contributing to the protection of essential infrastructures and the maintenance of operational continuity in critical contexts.

References

[1]. Modarres, M. (2016). Risk Analysis in Engineering: Techniques, Tools, and Trends. CRC Press.

[2]. Rausand, M. & Høyland, A. (2004). System Reliability Theory: Models, Statistical Methods, and Applications. Wiley-Interscience.

[3]. Kuo, W. & Zhu, D. (2013). Fault Diagnosis and Fault-Tolerant Control and Guidance for Aerospace Vehicles. Springer.

[4]. Kim, S. & Son, J. (2013). Fault Tree Analysis of Cybersecurity Attacks in Industrial Control Systems. International Journal of Security and Its Applications, 7(5), 65-76.

[5]. Bishop, M. (2003). Computer Security: Art and Science. Addison-Wesley.

[6]. ISO/IEC 27001:2013. Information Security Management Systems - Requirements. International Organization for Standardization.

[7]. Ross, R. & McEvilley, M. (2016). Systems Security Engineering: Considerations for a Multidisciplinary Approach in the Engineering of Trustworthy Secure Systems. NIST Special Publication 800-160.

[8]. Higham, D. J. & Higham, N. J. (2016). MATLAB Guide (3rd ed.). SIAM.- Manual practic pentru programare și simulare în Matlab, incluzând metode Monte Carlo.

[9]. MathWorks Documentation. (n.d.). Fault Tree Analysis. Retrieved from https://www.mathworks.com/help/reliability/fault-tree-analysis.html.

[10]. Sham Tickoo. (2022). PCB Design Using OrCAD Capture and PCB Editor. CADCIM Technologies. ISBN: 978-1640571470.

[11]. Fishman, G. S. (1996). Monte Carlo: Concepts, Algorithms, and Applications. Springer.

[12]. Rubinstein, R. Y., & Kroese, D. P. (2016). Simulation and the Monte Carlo Method (3rd ed.). Wiley.

[13]. Pan, K., Liu, H., Gou, X., Huang, R., Ye, D., Wang, H., Glowacz, A. & Kong, J. (2022). Towards a Systematic Description of Fault Tree Analysis Studies Using Informetric Mapping. Sustainability, 14(18), 11430.

Securing the Future: Cybersecurity Challenges in Wearable Devices

Daniela NAIPEANU
Faculty of Electronics, Telecommunications and Information Technology,
National University of Science and Technology POLITEHNICA Bucharest, Romania
naipeanu_daniela@yahoo.com

Abstract

Wearable devices, such as smartwatches, fitness trackers and medical monitors, have become essential to our day-to-day lives. They offer incredible benefits, making it easier to track our health and stay connected to the world. This kind of technology comes with a lot of advantages, as it simplifies different tasks people do daily. However, this convenience comes with significant cybersecurity risks. Because of the collected data, cybercriminals are starting to target the devices. This article will present the challenges faced to keep wearable technology secure. The paper examines the multiple cyber threats that the devices encounter, like malware, phishing and vulnerabilities in wireless connectivity. Also, the paper presents a software framework that is specifically built to offer uninterrupted security monitoring for wearable devices. The platform employs a complete strategy to tackle security challenges, encompassing intrusion detection and prevention, efficient vulnerability management, and prompt security patching.

Index terms: cybersecurity, cyberthreats, software, vulnerability, wearable devices

1. Introduction

Wearable technology is any kind of electronic device designed to be worn on the user's body. Such devices can take many different forms, including jewelry, accessories, medical devices, and clothing or elements of clothing. The rapid development and market introduction of wearable gadgets sometimes leads to the neglect of security and risk management. With the rapid increase in the adoption of wearable gadgets, their vulnerabilities are also expanding, leading to heightened concerns regarding security and privacy issues. The data gathered by wearable gadgets, if misused, has the potential to inflict greater harm than data obtained from smartphones and other devices [1].

In the last years, these devices have created another cybersecurity attack sector, which facilitates the data confidentiality and integrity of users and organizations [2]. The level of risk posed by technology increases as it becomes increasingly personalized to specific users [3]. The issue concerning wearable technologies lies in their security and privacy aspects, mostly owing to the absence of authentication, authorization, and unsafe methods of information transfer [1]. Organizations that permit the use of wearable devices in the workplace may not possess a comprehensive understanding of the various security vulnerabilities that these devices can introduce. Many enterprises, including major organizations with established security protocols, do not see these devices as potential risks to network safety and security. Inadequate management of wearable devices' assessment presents an escalating security hazard [4].

The great demand for wearable gadgets comes from their capacity to provide convenient and real-time delivery of important functionality [3]. These devices are equipped with sensors that facilitate the gathering of extensive volumes of data. The primary sensors include accelerometers, gyroscopes, Global Positioning System (GPS), acoustics, and voice detection [3]. As these integrated sensors collect personally identifiable information (PII), people and organizations become susceptible to vulnerabilities and cyber-attacks. As wearable gadgets become more popular, producers and developers seem to prioritize improving design aesthetics and power consumption over security features and dynamics [3]. As can be seen in the figure below, in recent years, many types of portable technologies have emerged.

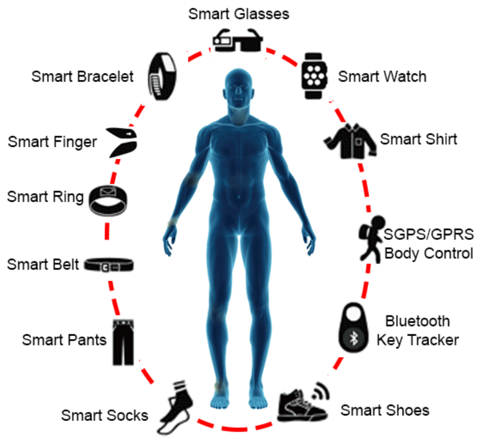


Fig. 1. Different types of wearable technology [13]

In the ever-evolving digital age, cybersecurity has become an increasingly pressing concern for both organizations and individuals alike. With emerging technologies and increasingly sophisticated threats, traditional security paradigms are no longer sufficient. Thus, the adoption of innovative and state-of-the-art approaches is necessary to counteract both current and future cyber threats.

One of these revolutionary paradigms is Zero Trust Security. Contrary to traditional models that focus on perimeter defense, Zero Trust assumes that no user, system, or network is inherently trustworthy. It mandates strict authentication and authorization based on multiple data points, such as user identity, location, device health, and the service or task being performed [5].

While Zero Trust focuses on controlled access, another major concern is resistance to future threats, such as those posed by quantum computers. Quantum-resistant algorithms, also known as post-quantum cryptography, are cryptographic algorithms that are designed to be secure even when attacked by quantum computers. Traditional cryptographic methods, such as RSA and ECC (Elliptic Curve Cryptography), rely on the difficulty of specific mathematical problems to ensure security. However, quantum computers have the potential to solve these problems much faster than classical computers, making these algorithms vulnerable [6].

Quantum-resistant algorithms, on the other hand, are built around mathematical problems that are thought to be difficult even for quantum computers to solve. Examples of cryptography include lattice-based, code-based, hash-based, and multivariate polynomial cryptography [6].

In addition to these innovations, biometric security has become increasingly widespread and sophisticated. Biometric security uses unique biological characteristics to authenticate people, providing a highly secure method of identity verification. It is based on fingerprints, iris patterns, facial recognition, voiceprints, and even behavioral characteristics such as typing patterns or gait. Unlike passwords or keys, biometric data is difficult to replicate, which improves security [7].

One significant benefit is its convenience. Users are not required to remember passwords or carry physical keys; their biological characteristics serve as their credentials. This simplicity encourages user adoption while lowering the risk of unauthorized access caused by forgotten passwords or lost keys [7].

2. Vulnerabilities and threats

Wearable devices often gather and transmit sensitive personal information, including health metrics, location data, and personal habits. The exposure of this data can have severe privacy implications; thus, this chapter presents a few of the primary cybersecurity vulnerabilities and threats associated with these types of devices.

Data Privacy and Security

Wearable devices frequently collect sensitive personal data, including health metrics and location information. If this data is not properly protected, it can be exposed to unauthorized parties, leading to severe privacy implications. Many wearables transmit data to smartphones or cloud services without adequate encryption, making it susceptible to interception during transmission [8].

Weak Authentication Mechanisms

Many wearable devices use weak or no authentication protocols, making it easy for unauthorized individuals to access the device and its data. Additionally, these devices often come with default passwords that users do not change, increasing the risk of unauthorized access [9].

Physical Security Risks

Wearables are easily lost or stolen, potentially granting attackers physical access to sensitive data. Physical access also allows attackers to tamper with the device's hardware or software, compromising its security [10].

Malicious Applications

The integration of third-party applications can introduce vulnerabilities if these apps are not properly vetted. Excessive permission requests by applications can lead to unnecessary exposure of personal data [11].

Interoperability and Connectivity

Wearables rely on various communication protocols, such as Bluetooth, which can be exploited through attacks like BlueBorne. Devices connected to Wi-Fi networks can also be targeted by network-based attacks, including man-in-the-middle attacks [11].

Insufficient End-User Awareness

Inadequate management and control over wearable devices, especially in corporate environments, can lead to unauthorized access to corporate networks and data. Bring Your Own Device (BYOD) policies complicate security efforts as wearables become part of personal devices brought to work [8].

3. A new solution

The growing incorporation of wearable electronics into daily routines has led to a significant rise in possible security vulnerabilities. The dangers encompass a wide range of risks, including data breaches and unauthorized access. Therefore, it is essential to have a strong and proactive security solution in place [12].This chapter presents a software framework that is specifically built to offer uninterrupted security monitoring for wearable devices. The platform employs a complete strategy to tackle security challenges, encompassing intrusion detection and prevention, efficient vulnerability management, and prompt security patching.

This platform relies on advanced machine learning algorithms and behavioral analysis methodologies. These features facilitate the platform's ability to quickly identify and address any threats or suspicious actions, hence guaranteeing the integrity and security of wearable devices. The subsequent sections will thoroughly examine the intricate data flow of the suggested solution, scrutinizing each component of the diagram to comprehend its function within the comprehensive security architecture.

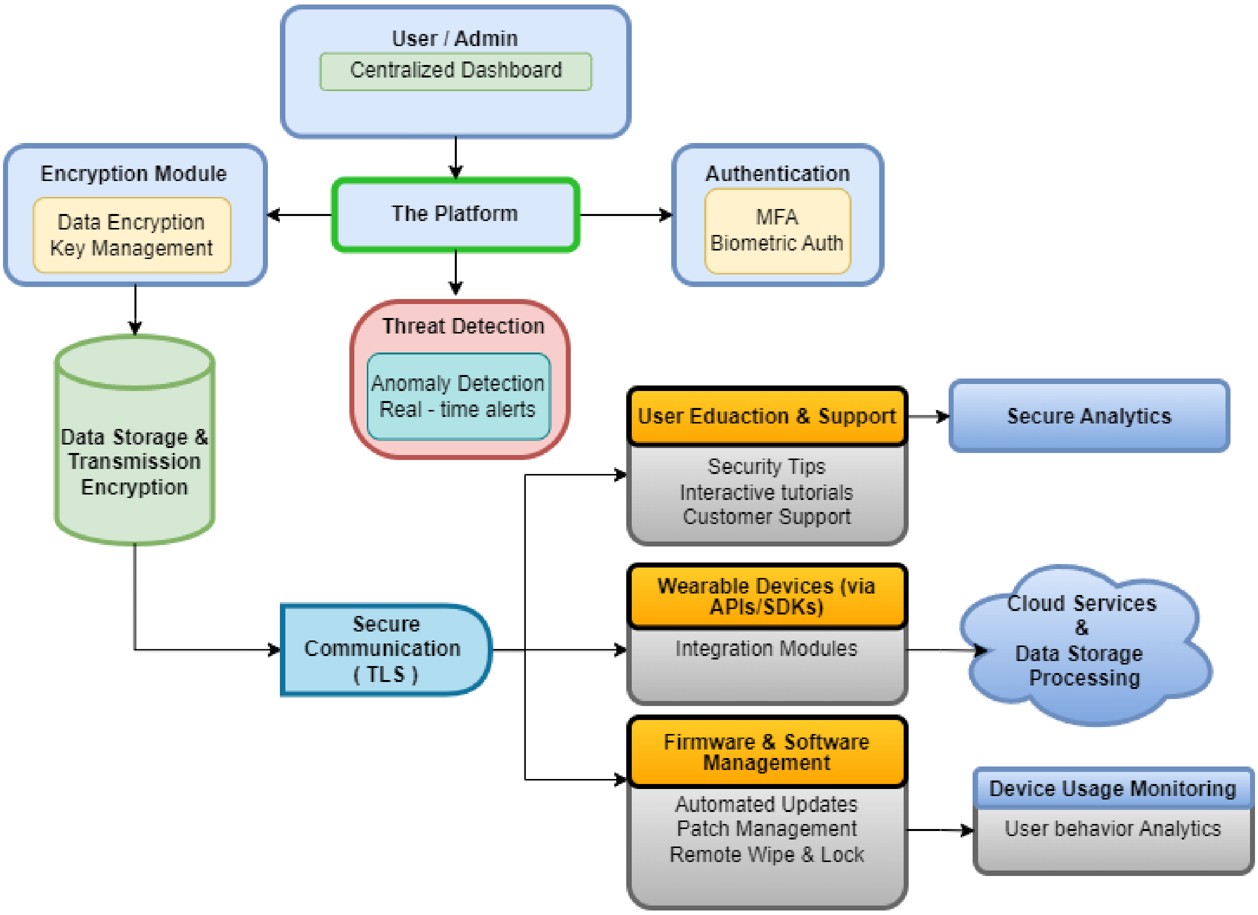


Fig. 2. Diagram of the software platform

The diagram illustrates a comprehensive security platform designed for continuous monitoring of wearable devices. This platform ensures robust protection through various integrated components and services. Every component, from user authentication to threat detection and real-time reaction, is essential for ensuring device security. The platform also incorporates functionalities for firmware and software administration, user instruction, and interaction with cloud services, guaranteeing thorough safeguarding and optimal functionality of wearable gadgets.

- **User / Admin access:** To start, the user or admin logs into the platform using their login details. The login details are verified by the platform's authentication service.
- **Centralized dashboard interaction:** After authentication, the user/administrator is redirected to the centralized dashboard, where they can view and interact with various security modules and features.
- **Encryption Module:** The user/administrator can access the encryption module from the dashboard to configure data encryption settings. Here, they can select encryption algorithms and manage encryption keys to ensure the security of stored and transmitted data.

- **Authentication Services:** Within the dashboard, the user/administrator configures authentication methods, such as Multi-Factor Authentication (MFA) and biometric authentication, to validate user identity and prevent unauthorized access.
- **Threat Detection and Response:** The platform continuously monitors wearable devices for abnormal behavior or security threats. Upon detecting a threat, the platform generates real-time alerts and notifies the user/administrator to take immediate remedial action.
- **Fimware & Software Management:** The user/administrator can initiate and manage firmware and software updates for wearable devices from the dashboard. The platform delivers and applies updates and patches to ensure optimal device performance and address any security vulnerabilities.
- **User Education & Support:** The platform provides user education resources, such as security tips and interactive tutorials, to enhance awareness and understanding of security measures. Additionally, customer support is available to address questions and requests related to security and platform usage.
- **Wearable Devices Integration:** Wearable devices integrate with the WearSecures platform through application programming interfaces (APIs) and software development kits (SDKs), enabling communication and data exchange between devices and the platform.
- **Cloud Services Interaction:** Data collected from wearable devices is stored and processed in cloud services for scalability and redundancy. The platform utilizes cloud services for real-time data processing, analysis, and reporting to identify threats and security trends.
- **Monitoring and Analytics:** The platform continuously monitors device usage, user behavior, and security events to identify potential risks and anomalies. Security analysis is performed to detect significant patterns, trends, and anomalies and provide useful information to the user/administrator for decision-making.

4. Implementation of the new methodology

The first step in creating a software platform for wearable device security is clearly defining the requirements. This process includes an initial analysis to determine the platform's functional and non-functional requirements. It is critical to consult with all stakeholders, including users and administrators, in order to understand the specific needs and establish detailed platform specifications.

Once the requirements have been established, the system design process begins. This includes developing an overall system architecture, which includes backend, frontend, and database decisions, as well as cloud service integration. Every major module, including encryption, authentication, monitoring, and firmware updates, must be clearly defined. It is also critical to create a user-friendly UI/UX for the centralized dashboard and associated interfaces.

Setting up the development and deployment environment, as well as version control tools like Git, is the first step toward effective development. Next, backend modules for authentication, encryption, and user and device management are created. In parallel, the frontend is being developed to create the user interface. API integration is a critical step in establishing communication between wearable devices and the platform. Security precautions are also taken to protect data and users.

Testing is an important step in ensuring the quality of the platform. It all starts with unit testing, which ensures that each module is correct. Then, integrated testing is performed to ensure that the module interactions work properly. Performance testing assesses the platform's behaviour under load, while security testing identifies and resolves any vulnerabilities.

Following launch, the platform should be constantly monitored for performance and security. Updates and enhancements are required to introduce new functionality and resolve potential issues. It is also critical to provide technical support to users and administrators in order to resolve issues and answer questions about how to use the platform.

5. Case Study: The 2016 Data Breach Involving 61 million Fitbit and Apple Users

In 2016, a significant data breach was uncovered, exposing the personal information of 61 million users of wearable fitness devices, including those from Fitbit and Apple, as well as other fitness-related applications and services. The breach highlighted serious concerns regarding privacy invasion, identity theft, and targeted phishing attacks [13].

Background: Wearable fitness devices and related applications have gained immense popularity over the years, allowing users to monitor their health and fitness levels. Companies like Fitbit and Apple have been at the forefront of this technological advancement, providing users with devices that track various health metrics, including steps taken, heart rate, sleep patterns, and more. However, the convenience and benefits of these devices come with inherent risks, especially concerning data security and privacy.

The Breach: In 2016, it was discovered that a database containing the personal information of 61 million users was inadvertently exposed online. This database primarily included users of Fitbit and Apple devices but also encompassed data from other fitness-related apps and services. The compromised data included sensitive information such as names, birthdates, weight, height, gender, location, and extensive activity logs. A third-party data management company was identified as the source of the breach. The company responsible for handling the data, had inadvertently made the database accessible online without proper security measures in place. This lapse in security protocols led to the unauthorized availability of millions of users' personal information [13] [14].

Impact and Concerns: The disclosure of such a vast amount of personal data raised several critical concerns:

- *Privacy Invasion*: The breach exposed intimate details of users' lives, including their physical activity and health metrics. Such information, in the wrong hands, could lead to severe privacy violations.
- *Identity Theft*: With access to personal identifiers such as names, birthdates, and location data, cybercriminals could potentially use this information to commit identity theft, opening fraudulent accounts, or conducting unauthorized transactions.
- *Targeted Phishing Attacks*: The detailed user information could be exploited to craft highly targeted phishing attacks. Cybercriminals could deceive users into divulging more sensitive information or infecting their devices with malware [13] [14].

Response and Mitigation: Upon discovery, immediate actions were taken to secure the exposed database and prevent further unauthorized access. The incident prompted a thorough investigation to understand the extent of the breach and to implement measures to prevent future occurrences. This included:

- *Securing the Database*: Ensuring that the database was no longer accessible to unauthorized individuals.
- *Reviewing Security Protocols*: Conducting a comprehensive review of the security practices of the third-party data management company and other associated entities.

- *User Notification*: Informing affected users about the breach and providing guidance on steps to protect their personal information, such as monitoring their accounts for suspicious activity and being cautious of phishing attempts [13] [14].

Lessons Learned: The 2016 data breach underscored the importance of robust data security measures, especially for companies handling sensitive personal information. Key takeaways from this incident include:

- *Vigilant Security Practices*: Organizations must ensure that all third-party partners adhere to strict security protocols to protect user data.
- *Regular Audits*: Conducting regular security audits and assessments can help identify and mitigate potential vulnerabilities.
- *User Education*: Educating users about the risks of data breaches and how to protect themselves can minimize the impact of such incidents [13] [14].

The 2016 data breach involving Fitbit and Apple users serves as a stark reminder of the vulnerabilities in our increasingly connected world. As technology continues to evolve, the importance of safeguarding personal information cannot be overstated. Companies must prioritize data security to protect their users and maintain their trust [13] [14].

6. Conclusion

The integration of wearable electronics into daily life has significantly increased security vulnerabilities, necessitating robust and proactive security solutions. The proposed software framework addresses these challenges through continuous monitoring, intrusion detection, and effective vulnerability management. By utilizing advanced machine learning algorithms and behavioral analysis, the platform ensures the integrity and security of wearable devices.

The platform's comprehensive design, featuring components like encryption modules, authentication services, and threat detection, provides thorough protection. Its user-friendly dashboard facilitates interaction, while cloud services and continuous monitoring enhance data processing and security. Implementation involves clear requirement definitions, system design, development, testing, and ongoing monitoring. This structured approach ensures the platform's effectiveness in protecting against security threats, ultimately safeguarding user data and maintaining trust in wearable technology. The 2016 breach exposing 61 million Fitbit and Apple users highlighted these vulnerabilities, revealing sensitive personal information due to lax security measures by a third-party company.

In summary, while wearable devices offer substantial benefits, ensuring their security and privacy is crucial. Organizations must implement comprehensive strategies to protect user data, maintain trust, and fully realize the potential of wearable technology.

References

[1]. A. J. Mills, R. T. Watson, L. Pitt and J. Kietzmann, "Wearing safe: Physical and informational security in the age of the wearable device," Business Horizons, vol. 59, no. 6, pp. 615-622, 2016.

[2]. F. Blow, Y.-H. Hu and M. A. Hoppa, "A Study on Vulnerabilities and Threats to Wearable Devices," The Colloquium for Information Systems Security Education, vol. 7, no. 1, 2020.

[3]. K. W. Ching and M. M. Singh, "Wearable Technology Devices Security and Privacy Vulnerability Analysis," International Journal of Network Security & Its Applications (IJNSA), vol. 8, no. 3, pp. 19-30, 2016.

[4]. A. G. Silva-Trujillo, M. J. González González, L. P. Rocha Pérez and L. J. García Villalba, "Cybersecurity Analysis of Wearable Devices: Smartwatches Passive Attack," Sensors, 2023.

[5]. S. Rauch, "What Is Zero-trust Security? A Handy Guide," Simplilearn, 26 Feb 2023. [Online]. Available: https://www.simplilearn.com/tutorials/cyber-security-tutorial/zero-trust-security.

[6]. "NIST to Standardize Encryption Algorithms That Can Resist Attack by Quantum Computers," NIST, 24 August 2023. [Online]. Available: https://www.nist.gov/news-events/news/2023/08/nist-standardize-encryption-algorithms-can-resist-attack-quantum-computers.

[7]. F.-Z. Marcos, "Biometric security technology," Aerospace and Electronic Systems Magazine, pp. 15-26, 2006.

[8]. S. Saleem, S. Ullah and K. S. Kwak, "A study of IEEE 802.15.2 Security Framework for Wireless Body Area Network," Sensors, pp. 1383-1395, 2011.

[9]. A. Bianchi and I. Oakley, "Wearable Authentication: Trends and Opportunities," IT-Information Technology, vol. 58, no. 5, pp. 225-262, 2016.

[10]. I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "A survey on sensor networks," IEEE Comm. Mag, vol. 40, pp. 102-114, 2002.

[11]. C. Karlof and D. Wagner, "Secure routing in wireless sensor networks," Proceedings of the 1st IEEE International Workshop on Sensor Network Protocols and Applications, pp. 113-127, 2003.

[12]. M. Syafrizal, S. R. Selamat and N. A. Zakaria, "Analysis of Cybersecurity Standard and Framework Components," International Journal of Communication Networks and Information Security (IJCNIS), vol. 12, no. 3, pp. 417-432, 2020.

[13]. H. Landi, "Fitbit, Apple user data exposed in breach impacting 61M fitness tracker records," Fierce Healthcare, 13 Sep 2021. [Online]. Available: https://www.fiercehealth care.com/digital-health/chutes-ladders-nyc-health-appoints-inaugural-decarbonization-officer-uber-health.

[14]. J. McKeon, "61M Fitbit, Apple Users Had Data Exposed in Wearable Device Data Breach," Intelligent Healthcare Media is a division of TechTarget, [Online]. Available: https://healthitsecurity.com/news/61m-fitbit-apple-users-had-data-exposed-in-wearable-device-data-breach.

[15]. H. A. Junqueira, "Different types of wearable technology," ResearchGate, [Online]. Available: https://www.researchgate.net/figure/Different-types-of-wearable-technology _fig5_322261039.

Public Attribution in Cyberspace: Symbolic Gesture or Strategic Weapon?

Mihai OLTEANU
National Defense University "Carol I", Bucharest, Romania
mihaiolteanu48@yahoo.com

Abstract

States have developed multiple defensive approaches, but deterrence has emerged as a central element of cyber strategy. Public attribution (the official act of publicly identifying the actor responsible for a cyber operation) has become a widely used instrument, particularly among NATO and EU members, despite its high political and technical costs. This paper examines the relationship between public attribution and deterrence, with reference to five recognized forms: punishment, denial, entanglement, norms, and association. Using recent data and selected cases, the analysis shows that public attributions can sometimes generate tangible outcomes (e.g., sanctions, indictments, or defensive improvements), while in other cases they produce little deterrent effect. Nonetheless, the value of public attribution should not be measured solely by utilitarian efficiency, but also through the principled lens of international law and normative signaling. Even when immediate effects are absent, attribution reaffirms international norms, imposes reputational costs, and prevents the normalization of hostile behavior in cyberspace.

Index terms: Advanced Persistent Threat, cybersecurity, deterrence, international law, public attribution

1. Introduction

The last two decades have witnessed a continuous growth in the number of cyberattacks, characterized by diverse technical features, varying levels of complexity and innovation, and a substantial number of victims worldwide. Notably, since 2013, the main area of interest of cyberattacks has shifted constantly, from operations focused on cloud computing (2013), Internet of Things (2014), cryptocurrency (2017), and the healthcare system (2019-2020), to campaigns that reflected the rise of Artificial Intelligence in automating and accelerating attacks [1], highlighting a connection between cyberattacks and emerging technologies.

While the overall landscape of cyberattacks has evolved rapidly in scope and focus, Advanced Persistent Threats (APT) campaigns represent a somewhat distinct phenomenon. Although these operations account for a smaller share of the total incidents compared to financially or ideologically motivated attacks, APT operations remain the most relevant threat in terms of strategic impact for both states and companies, as their outcome is likely to fundamentally affect the activity of these entities. APT campaigns have evolved their tools but have also remained stable in terms of core characteristics (e.g., long-time persistence, highly-tailored targeting of specific domains, and significant funding) and modus operandi [2].

In dealing with APT campaigns, multiple techniques and procedures have been developed by state agencies and private companies, focused on each level of decision: strategic, operational, and tactical. Tactically, the defensive mechanisms evolved by integrating tools such as Intrusion Detection

Systems and Endpoint Detection and Response systems, while operationally, the use of advanced tools for anomaly detection and behavioral analysis strengthened the security of the systems [3]. Still, strategically, the practice of deterrence in cyberspace has remained a continuous component in handling massive APT campaigns. Such an approach is rather specific to governments, as private companies are more oriented towards financial gains, instead of strategic statements and international sanctions.

The practice of deterrence is the result of a set of strategic approaches that ultimately aim to discourage the threat actor from conducting campaigns against a target or a group of targets [4]. The effectiveness of traditional deterrence strategies is directly dependent on technical arguments that outline the identity of the author of a cyberattack, a process generally referred to as attribution. Naturally, the aim of different states to impose various deterrence measures requires, in most cases, a previous attribution of a cyber campaign, which usually becomes public if enough data is gathered, to serve as a foundation for additional diplomatic, economic, or technical measures. By officially identifying and exposing the responsible actor, governments attempt to impose reputational costs, reduce the adversary's room for plausible deniability, and mobilize international support against common threats [5].

During the last decades, public attribution of cyberattacks has become a traditional tool in dealing with cyberattacks, with examples dating back to 2007 (Estonia to Russia [6]), 2010 (Google to China [7]), and 2014 (USA to North Korea [8]). However, there are constant debates surrounding the efficiency of such approaches, considering the various reactions of the threat actors after the public attribution is conducted.

This paper aims to analyze multiple approaches to public attribution of cyberattacks, looking to provide arguments for and against this process, from a utilitarian perspective. In doing so, several deterrence mechanisms will be evaluated, along with some examples of public attributions and the outcomes they generated. Considering that this approach is rather specific to governments, as it is a predecessor of other diplomatic or technical actions, in accordance with international law, the public attributions performed by private companies will not be discussed in this paper.

2. Deterrence in cyberspace

The persistence of APT campaigns, as well as the high stakes implied by such operations, oblige governments to address the matter by exploiting a mixture of decisions on each of the three levels. Therefore, while tactical and operational approaches mainly resemble financial investments, strategic decisions are usually focused on various deterrence mechanisms, shaped to generate expected behaviors from threat actors.

Joseph S. Nye described deterrence as a set of means that dissuade adversaries from conducting operations by making them believe that the expected benefits are exceeded by the most probable costs [9]. Deterrence is only useful if, psychologically, all the actors involved calibrate their perceptions regarding the costs and benefits. The most reliable example of deterrence that serves such a definition is the one regarding the nuclear threat, which implies a set of outcomes that makes any attempt at such conflict seem unworthy.

The author further described four approaches to this strategy: (1) deterrence by punishment, where the retaliation is likely to produce unwanted or unacceptable damage to the initiator of the attack, thus making the initial offensive unfit; (2) deterrence by denial, essentially consisting of building the required capabilities to prevent or limit the damage, therefore convincing hostile operations that their chances of success are reduced; (3) deterrence by entanglement, which implies that the aggressor and the victim are so well interconnected that an attack against the latter is likely to produce damages also against the former, concluding to a financially unfeasible situation and (4) deterrence by norms, implying a reputational damage to the aggressor, as the break of international rules may affect its soft power capabilities in the medium and long run [9]. The variety of deterrence strategies implies that not each one of them works for any type of aggressor, as some might be more preoccupied with their international reputation (e.g., China [10]), and some with the retaliation, given that their cybersecurity level is rather low.

N.J. Ryan provides a complementary perspective by adding a fifth mechanism, deterrence by association, a term initially used by Paul Cornish. His theory essentially states that the risk of being connected with the outcomes of a significant cyberattack through public attribution is a setback both for governments, which may face different types of strategic consequences, such as public condemnation, and for individuals, who could be indicted [11]. Such cases have been constantly identified, with several American public attributions directly naming some individuals who conducted cyber operations and are being legally pursued for crimes such as conspiracy to damage protected computers, economic espionage, identity theft, or conspiracy to commit computer intrusions [12] [13] [14].

Out of the five types of approaches, deterrence by denial and deterrence by entanglement are indifferent to the existence of an attribution, be it public or classified. However, it could be argued that these two categories also pose the lowest level of deterrence, as the first one is purely based on a financial analysis, concomitantly with the fact that some of the most evolved defensive mechanisms have eventually been breached. The second one is also based on an economic calculus, and could easily be overcome if the benefits obtained in a sector are more relevant than the damage produced by the existence of an interconnectivity in another. For instance, if compromising an interconnected portion of an energy network would generate immediate military advantages, such as control over essential areas, the cyber operations may seem to produce sufficient benefits to justify their existence and the final damage.

The other three deterrence mechanisms are directly linked to the preexistence of an attribution process, generally joined by at least a small degree of declassified details, that would publicly and internationally justify retaliation, normative punishment, or reputational damage through association with the authors of the operation. Therefore, a link between publicly attributing a cyber operation and most of the existing deterrence measures is evident, as any statement or action has to be based on some degree of evidence. Attribution becomes the enabling condition, since without a credible identification of the attacker, no measure can be effectively targeted. At the same time, the act of public attribution in itself functions as a deterrent by exposing the adversary, reducing plausible deniability, and signaling collective resolve to broader audiences.

3. Public attribution of cyberattacks

The public attribution of a cyberattack refers to the official act of publicly naming the actor deemed responsible for an operation, often based on a combination of technical forensics, intelligence assessments, and strategic considerations. The act of publicly attributing a cyberattack can take place in various ways, mainly ranging from public statements made by governments or institutions to joint communiqués performed by a set of international partners [15].

Generally, the public attribution of a cyberattack requires a consistent set of technical, financial, and political resources. Therefore, inherently, the investment of such an amount of resources has to be justified by an outcome that produces proportional advantages [16]. In this sense, public attribution is rarely employed for opportunistic or cybercrime activities, but rather for APT campaigns, which are typically state-sponsored, long-term, and strategically significant. The persistence, geopolitical relevance, and high-value targets of APTs make them the most likely category of operations to warrant the political and strategic effort of an attribution [17].

The European repository for Cyber Incidents distinguishes between five types of attributions performed by different entities [18]:

- Technical attribution, meaning the assignment of technical responsibility for a cyber incident by analyzing the attackers' tactics, techniques, and procedures to infer the likely origin of the operation. Such an attribution is rather specific to private cybersecurity companies, which publish such technical reports regularly, as part of their usual activity [19].
- Unofficial Policy Attribution, the process of attribution of responsibility for a cyber incident that may include statements or leaks from public officials or senior civil servants reported in the media.
- Official Policy Attribution, based on public statements, press releases, or speeches by government officials or senior civil servants. Unlike technical or legal attribution, this form of attribution generally presents less direct evidence and may emphasize political or strategic considerations. This is the most common form of public attributions performed by states.
- Legal Attribution, the process aimed at prosecution or the formal declaration of a violation of legal norms. It typically involves presenting verifiable evidence and supporting documentation to substantiate the claim. Such examples include extensive and complex technological investigations that typically end with naming not only states or agencies, but also individuals.
- Self-Attribution, a process conducted by the perpetrator, such as hacktivists or cybercriminals, often via social media or other public channels. Motivations for such disclosures may include intimidating the victim, signaling capabilities, or increasing pressure in the context of ransomware or extortion operations [18].

Considering that this study is focused on public attribution as a component of deterrence, we will mostly focus on the Official Policy Attribution, as it is the most common outcome of the efforts of a government towards naming a foreign perpetrator that managed to compromise its networks and produce damage.

The public attribution itself is a complex process, difficult to correctly and completely argue, given the nature of cyberspace, known for its volatility and high degree of anonymity. Compared with traditional means of deterrence, the public attribution of cyberattacks faces additional challenges. For instance, when discussing deterrence in the nuclear field, the limited number of states that possess such an arsenal is sufficient to limit the list of possible perpetrators [9]. Other attribution processes focused on the military area could be correctly argued through tangible evidence, such as pictures or recordings, while this is not the case for cyberattacks.

Given the high investments required to perform public attribution in cyberspace, as well as the difficulty in gaining strategic advantages through deterrence mechanisms, a further analysis of the efficiency of such processes is required.

4. Evaluating the public attribution

According to data published by the European Repository of Cyber Incidents on September 26, 2025, there have been 1,617 public attributions of all five types worldwide since 2022, with 704 of these being self-attributions. Therefore, there were 902 technical, legal, unofficial, and official policy attributions during the last five years, out of which:

- 215 (approx. 23.8%) have pointed to Russian private or governmental entities as being the authors of the attacks.
- 167 (approx. 18.5%) have indicated Chinese private or governmental entities as being the authors of the attacks.
- 107 (approx. 11.8%) have pointed to North Korean entities as being the authors of the attacks.

- 91 (approx. 10%) have indicated Iranian private or governmental entities as being the author of the attacks [18].

Therefore, 64.1% of all public attributions conducted over the past three years identified traditional NATO and EU adversaries as the perpetrators of cyberattacks. This distribution suggests that public attribution is regularly used as a strategic instrument against adversaries in cyberspace by EU and NATO members. By attributing cyberattacks to adversarial states, NATO and EU members reinforce a narrative of external threats, thereby justifying defensive policies, strengthening alliances, and deterring future aggression [20].

More specifically, since 2022, out of the total of 902 public attributions, the most have been initiated from:

- US, 391 attributions (approx. 43.3%).
- EU states, 151 attributions (approx. 16.7%).
- Ukraine, 57 attributions (approx. 6.3%).
- UK, 42 attributions (approx. 4.6%) [18].

The fact that the majority of public attributions (64.6%) are issued by NATO and EU states highlights important dynamics. These data demonstrate the political will and strategic approach of the Western states, as the attribution is not solely a technical act but also a political choice, and NATO/EU members have consistently employed it as a tool of deterrence and signaling [21].

However, public attribution has also been employed by adversarial states through different approaches. On one hand, China has made limited use of this tool, issuing only 13 public attributions over the past three years. On the other hand, Russian entities have conducted 157 public attributions, the majority of which (119 cases, 75.7%) were self-attributions, typically made by hacktivist groups that publicly claimed responsibility for cyberattacks [18].

The asymmetry between NATO/EU members and Russian attribution practices is significant. On the Western side, the majority of public attributions target Russia, portraying it as the primary source of malicious cyber activity and reinforcing its image as a systemic threat. This reflects a deliberate strategic use of attribution as a tool of deterrence. By contrast, most Russian attributions are not directed but are instead self-proclaimed, with hacktivist groups claiming responsibility for attacks. In this sense, public attribution in the Russian context operates less as a mechanism of assigning external responsibility and more as an act of demonstrating its capabilities.

Ukraine represents a hybrid case, which seems to combine elements of both Western and Russian practices, as it has issued 120 public attributions, of which more than half (63, approx. 52.5%) are self-attributions. On one hand, Ukraine engages in outward-facing attribution, using public attribution as a strategic tool for naming and shaming. On the other hand, the significant proportion of self-attributions (often claimed by Ukrainian hacktivists or even governmental institutions) reflects a strategy where self-proclaimed operations are used to project resilience, demonstrate offensive capability, and, possibly, sustain domestic and international morale during wartime.

Still, the efficiency of public attribution is debatable. There have been cases that illustrated several outcomes produced by such initiatives, indicating that the strategic use of public attribution has been a relevant tool. In 2021, a worldwide campaign dubbed HAFNIUM revealed the exploitation of several zero-day vulnerabilities in Microsoft Exchange servers, affecting a variety of industries, with an estimated number of more than 30.000 in the US alone, and hundreds of thousands globally [22]. While Microsoft released several statements on this matter, along with a set of patches to prevent further exploitation of the zero-day vulnerabilities [23], there was a continuous growth in the number of victims. In July 2021, the UK, alongside international partners, issued an official public attribution holding Chinese state-backed actors responsible for the exploitation of Microsoft Exchange Server vulnerabilities. The UK's National Cyber Security Centre assessed with high confidence that the group HAFNIUM was almost certainly affiliated with the Chinese state and that the operation was likely intended to enable large-scale cyber espionage. The UK also called on China to adhere to its 2015 G20 commitment not to engage in cyber-enabled intellectual property theft [24].

While the attribution itself did not lead to deterrence by retaliation, by entanglement, or by denial (as, by the end of March 2021, 92% of the potential victims patched the vulnerabilities [24]), it posed, however, several elements indicating deterrence by association and by normative taboos, proving a relevant level of efficiency. Precisely, this initiative functioned as deterrence by association, considering that China felt obliged to issue a response to the public attribution, in which it firmly denied and called these claims unreasonable [25]. Moreover, deterrence by normative taboos functioned by exposing and denouncing state behavior that did not follow established cyber norms (particularly focused on large-scale espionage and the theft of intellectual property). Even without immediate punitive consequences, the act of public shaming contributes to long-term normative pressure against such conduct. Moreover, the official response of the Chinese authorities pointed out that the country is committed to these norms by opposing and condemning any form of cybercrime [25].

Similarly, other public attributions managed to produce some forms of deterrence, such as the one initiated by the US in March 2024, naming and sanctioning a group of defendants that were linked to the cyber operations conducted by APT31, a threat actor publicly attributed to the Chinese Ministry of State Security [26]. Similarly, France's individual public attribution of APT28's operations to the Russian military intelligence service marked a notable precedent, being the first such attribution issued independently by France against Russia [27]. It also produced important outcomes aiming to improve deterrence by denial by using this public attribution as an argument to update its strategic view (that, unlike other versions, explicitly points out Russia as the primary threat) [28] and to task the ANSSI (Agence nationale de la sécurité des systèmes d'information) to strengthen the national cybersecurity [29].

On the other side, there are examples of public attributions that did not produce an important outcome in terms of deterrence. Simply evaluating the high number of public attributions conducted during the last three years points out that the mechanism has become a common tool used mostly by EU or NATO states to publicly shame a threat actor. For instance, while APT SANDWORM has been constantly blamed and shamed for conducting cyberattacks against European infrastructures [30] [31], its operations continued, managing to compromise a consistent number of victims. This suggests that while public attribution has become a widely and often used instrument, it does not inherently produce quantifiable strategic outcomes of any kind, as adversaries react differently to similar incentives. However, from a tactical perspective, the public attribution of an ongoing cyberattack can function as a form of deterrence by denial, as it signals that the intrusion has been detected and forces the perpetrator to change and adapt its techniques [17]. This, in turn, imposes additional costs in terms of time, manpower, and resources.

5. Conclusions

The analysis of public attribution of cyberattacks demonstrates that this mechanism has become a common and visible instrument in the cyber field, particularly among NATO and EU members. However, its efficiency as a tool of deterrence is not uniform. Certain cases, such as the public attribution of HAFNIUM and APT31, illustrate that public attributions can generate strategic outcomes (e.g., international accusations, improvement of defensive capabilities) or even tangible ones (e.g., indictments). In contrast, other instances, such as the repeated attributions against APT SANDWORM, might indicate little effect in preventing further hostile cyber operations, as the perpetrators continued their activities.

Still, this variation of outcomes highlights that public attribution should not be assessed solely through the lens of utilitarian efficiency. While deterrence by denial, punishment, norms, retaliation, or entanglement may sometimes be achieved, the act of attribution carries value even in the absence of immediate strategic outcomes. Public attribution reaffirms international principles, signals consistency in policy, and serves as a normative act of designating unacceptable behavior [21]. In this sense, it functions as much as a political and moral statement as it does as a practical deterrent. Moreover, the failure to attribute would risk normalizing hostile cyber operations and signaling acceptance of such behavior.

Furthermore, the distribution of public attributions in recent years underlines the political dimension of this practice. The majority of attributions have been performed recently by a relatively small group of states (mainly the US, EU members, the UK, and Ukraine). Russia and China, by contrast, employ this tool only rarely and in distinctive ways. Chinese authorities have issued very few official public attributions, while Russian practices are dominated by self-attributions claimed by hacktivist groups, often with the purpose of intimidation, propaganda, or boasting of capabilities. Russia's decision to use self-attribution less as a mechanism of external accountability and more as a performative act of demonstration highlights that public attribution can sometimes become a form of strategic communication [21], shaping perceptions of legitimacy and responsibility in cyberspace. Taken together, these practices suggest that Russia uses indifference to public attribution as a strategic posture, signaling resilience and trying to diminish the credibility of attribution as a deterrent tool. Still, at a minimum, public attribution signals that actions presumed to remain hidden have been uncovered, obliging the actor to reassess its operational security and tactical approach, hence requiring investments [17].

Ukraine, positioned in the middle of these approaches, has combined Western-style naming and shaming with a significant proportion of self-attributions, particularly during wartime. Its objective might be to signal resilience, deterrence, and offensive capacity to both domestic and international audiences.

Finally, while not all public attributions are effective in producing direct deterrent outcomes, their broader value lies in their cumulative and symbolic impact. Repeatedly attributing hostile cyber operations to adversaries contributes to long-term reputational costs and strengthens normative taboos. Even in cases where immediate deterrence is absent, attribution remains necessary to uphold international norms, maintain consistency, and avoid the erosion of standards in cyberspace. The act of attribution, therefore, should be understood not only as a utilitarian tool but also as a principled practice.

6. Discussions

This paper does not argue against the practice of publicly attributing cyber operations; on the contrary, it advocates for its continuation, as it reinforces existing norms, highlights the responsible actor, and signals unacceptable behavior in cyberspace. The main argument that can be extracted from this paper is that public attribution should not be assessed solely in terms of efficiency, but rather through the principled lens of international law, which obliges states to act against perpetrators even when such actions do not produce measurable effects, but serve to uphold valuable international norms.

However, a quantitative analysis assessing the effectiveness of public attributions in supporting each of the five forms of deterrence could provide a conclusive perspective on this discussion. Moreover, such an analysis should differentiate between types of aggressors, as their responses to public attribution are likely to vary depending on domestic dynamics and international strategies or behaviors.

Nevertheless, regardless of the outcomes of such assessments, the principle of publicly attributing a cyberattack remains valid and should not be judged solely through the lens of immediate efficiency. Therefore, public attribution is rather a symbolic gesture with strategic value.

References

[1]. M. Elgan, "A decade of global cyberattacks, and where they left us," IBM, 9 July 2024. [Online]. Available: https://www.ibm.com/think/insights/decade-global-cyberattacks-where-they-left-us. [Accessed 2 September 2025].

[2]. A. Alageel and S. Maffeis, "Investigation of Advanced Persistent Threats Network-based Tactics, Techniques and Procedures," Computer Science, 2025.

[3]. N. Ibrahim, N. Rajalakshmi and K. Hammadeh, "Exploration of Defensive Strategies, Detection Mechanisms, and Response Tactics against Advanced Persistent Threats APTs," Nanotechnology Perceptions, vol. 20, no. 4, pp. 439-455, 2024.

[4]. R. L. Kugler, "Deterrence of cyber attacks," Cyberpower and national security, vol. 320, pp. 309-340, 2009.

[5]. D. Tran, "The Law of Attribution: Rules for Attribution the Source of a Cyber-Attack," Yale JL & Tech, vol. 20, p. 376, 2018.

[6]. A. Schmidt, "The Estonian Cyberattacks," in The fierce domain - conflicts in cyberspace 1986-2012, Washington, D.C., Atlantic Council, 2013.

[7]. F. Nasir and R. Sohail, "State, Surveillance, and Cyberspace: An Intelligence Analysis of Operation Aurora," Social Science Review Archives, vol. 3, no. 2, 2025.

[8]. S. Goel, "How improved attribution in cyber warfare can help de-escalate cyber arms race," Connections , vol. 19, no. 1, pp. 87-95, 2020.

[9]. J. S. Nye, "Deterrence and Dissuasion in Cyberspace," International Security, vol. 41, no. 3, p. 44-71, 2017.

[10]. C. Lu and L. Zhang, "A Chinese Perspective on Public Cyber Attribution," China Quarterly of International Strategic Studies, vol. 8, no. 1, pp. 61-77, 2022.

[11]. N. J. Ryan, "Five Kinds of Cyber Deterrence," Philosophy & Technology, vol. 31, p. 331-338, 2018.

[12]. FBI, "APT 41 GROUP," 3 September 2020. [Online]. Available: https://www.fbi.gov/wanted/cyber/apt-41-group. [Accessed 21 September 2025].

[13]. FBI, "APT 10 GROUP," 7 December 2018. [Online]. Available: https://www.fbi.gov/wanted/cyber/apt-10-group. [Accessed 21 September 2025].

[14]. FBI, "APT 40 CYBER ESPIONAGE ACTIVITIES," 24 November 2020. [Online]. Available: https://www.fbi.gov/wanted/cyber/apt-40-cyber-espionage-activities. [Accessed 21 September 2025].

[15]. F. Egloff and A. Wenger, "Public Attribution of Cyber Incidents," CSS Analyses in Security Policy, vol. 244, 2019.

[16]. J. R. Lindsay, "Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack," Journal of Cybersecurity, vol. 1, no. 1, pp. 53-67, 2015.

[17]. F. J. Egloff and M. Smeets, "Publicly attributing cyber attacks: a framework," Journal of Strategic Studies, vol. 46, no. 3, pp. 502-533, 2021.

[18]. European Repository of Cyber Incidents, "Attribution Tracker," [Online]. Available: https://eurepoc.eu/attribution-tracker/. [Accessed 20 September 2025].

[19]. S. Romanosky and B. Boudreaux, "Private-Sector Attribution of Cyber Incidents: Benefits and Risks to the U.S. Government," International Journal of Intelligence and CounterIntelligence, vol. 34, no. 3, pp. 463-493, 2021.

[20]. G. Baram, "Cyber Diplomacy through Official Public Attribution: Paving the Way for Global Norms," International Studies Perspectives, 2024.

[21]. R. Gottemoeller, K. Hedgecock, J. Magula and P. Poast, "Engaging with emerged and emerging domains: cyber, space, and technology in the 2022 NATO strategic concept," Defence Studies, vol. 22, no. 3, pp. 516-524, 2022.

[22]. M. Grigutytė, "Microsoft Exchange zero-day vulnerability: What do you need to know?," NordVPN, 7 April 2024. [Online]. Available: https://nordvpn.com/blog/microsoft-exchange-exploits/. [Accessed 26 September 2025].

[23]. Microsoft 365 Security; Microsoft Threat Intelligence, "HAFNIUM targeting Exchange Servers with 0-day exploits," 2 March 2021. [Online]. Available: https://www.microsoft.com/en-us/security/blog/2021/03/02/hafnium-targeting-exchange-servers/. [Accessed 26 September 2025].

[24]. Foreign, Commonwealth & Development Office; National Cyber Security Centre; The Rt Hon Dominic Raab, "UK and allies hold Chinese state responsible for a pervasive pattern of hacking," 19 July 2021. [Online]. Available: https://www.gov.uk/government/news/uk-and-allies-hold-chinese-state-responsible-for-a-pervasive-pattern-of-hacking. [Accessed 26 September 2025].

[25]. BBC, "China says Microsoft hacking accusations fabricated by US and allies," 20 July 2021. [Online]. Available: https://www.bbc.com/news/world-asia-china-57898147. [Accessed 26 September 2025].

[26]. Department of Justice, "Seven Hackers Associated with Chinese Government Charged with Computer Intrusions Targeting Perceived Critics of China and U.S. Businesses and Politicians," 25 March 2024. [Online]. Available: https://www.justice.gov/archives/opa/pr/seven-hackers-associated-chinese-government-charged-computer-intrusions-targeting-perceived. [Accessed 26 September 2025].

[27]. Ministère de l'Europe et des Affaires étrangères, "Russia - Attribution of cyber attacks on France to the Russian military intelligence service (APT28) (April 29th, 2025)," 29 April 2025. [Online]. Available: https://www.diplomatie.gouv.fr/en/country-files/russia/news/2025/article/russia-attribution-of-cyber-attacks-on-france-to-the-russian-military. [Accessed 26 September 2025].

[28]. Secrétariat général de la défense et de la sécurité nationale, "National Strategic Review," 2025. [Online]. Available: https://www.sgdsn.gouv.fr/files/files/Publications/20250713_NP_SGDSN_RNS2025_EN_0.pdf. [Accessed 26 September 2025].

[29]. P. Madhiraju, "France says Russia-backed APT28 hackers targeted Olympic networks and government systems," 1 May 2025. [Online]. Available: https://business-news-today.com/france-says-russia-backed-apt28-hackers-targeted-olympic-networks-and-government-systems/. [Accessed 26 September 2025].

[30]. A. Waldman, "Mandiant upgrades Sandworm to APT44 due to increasing threat," 17 April 2024. [Online]. Available: https://www.techtarget.com/searchsecurity/news/366581178/Mandiant-upgrades-Sandworm-to-APT44-due-to-increasing-threat. [Accessed 26 September 2025].

[31]. Cybersecurity and Infrastructure Security Agency, "New Sandworm Malware Cyclops Blink Replaces VPNFilter," 23 February 2022. [Online]. Available: https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-054a. [Accessed 26 September 2025].

The Web Ecosystem Between Vulnerability and Resilience: The Case of Polyfill.io

Adelaida STĂNCIULESCU
Bucharest Court, Bucharest, Romania
adelaida.stanciulescu@gmail.com

Abstract

This report addresses the Polyfill.io security incident, analyzed as a case study to demonstrate the impact that poor management of external dependencies can have on the resilience of the modern web ecosystem. The Polyfill.io case represents one of the most extensive external dependency compromises in recent history, affecting over 100,000 websites through sophisticated mechanisms for injecting malicious code via compromised CDN infrastructure. The analysis reveals how a supposedly harmless JavaScript library, used for cross-browser compatibility, was hijacked and exploited as a global attack vector. The study investigates the mechanisms by which the polyfill.io domain was taken over and used for the conditional distribution of malicious code. Through comparative analysis with other major supply incidents chain (SolarWinds, Log4Shell, XZ Utils), the paper identifies the unique features of the Polyfill.io case - including the passive nature of the compromise, the almost instantaneous speed of propagation, and the unprecedented diversity of victims. The results of the analysis reveal the importance of implementing strengthened security measures for managing external dependencies, such as systematically verifying the integrity of resources, enforcing content security policies, and continuously monitoring ownership changes within open- source projects.

Index terms: CDN compromises, external dependencies, open-source security, Polyfill.io, supply chain attacks

1. Introduction

1.1. The general context of supply chain attacks chain in the contemporary web ecosystem

The contemporary web ecosystem has undergone a significant structural transition from traditional monolithic applications to distributed architectures based on microservices and extensive integration of external dependencies. This evolution has generated considerable advantages in terms of scalability and speed of software development, but has also introduced new levels of complexity and vulnerability from a cybersecurity perspective [14], [15], [16].

The contemporary web ecosystem is characterized by a complex interconnection of multiple components: web applications, servers, cloud services, content delivery networks (CDNs), and a multitude of APIs. In modern software development, code reuse through external libraries and open-source services is a standardized practice, which accelerates innovation and reduces implementation costs [14], [15].

However, this interdependence raises security concerns: compromising a single external component can generate a domino effect, propagating the vulnerability to tens or hundreds of thousands of applications. Recent examples (SolarWinds, Log4j, Polyfill.io, XZ Utils) have demonstrated that the modern web ecosystem is not just a technological assembly, but also a fragile chain of trust [9], [10], [14], [15].

Software supply chain attacks are one of the most insidious and effective forms of cyberattacks. Unlike traditional attacks that directly target an application or website, these attacks exploit vulnerabilities in the ecosystem of dependencies.

Two main types of attacks relevant to the present study are identified in the specialized literature:

- **Attacks on open- source dependencies** - consist of compromising a popular library published on platforms such as *npm* (Node Package Manager). A notorious example is the "event- stream" incident, where a widely used package was hijacked to insert malicious code.
- **Attacks on distribution infrastructure** - consist of compromising or acquiring content delivery services (especially CDNs), through which code reaches end-user browsers directly. This is exactly the category illustrated by the *Polyfill.io incident*.

The Polyfill.io case was particularly serious because the service was deeply embedded in the code of a significant number of websites and had a high level of trust from the developer community. This positioning made it an ideal vector for a supply -side attack. chain on a global scale.

1.2. External dependencies: definitions and classifications

In a strict sense, an external dependency is any third-party software component or service integrated into an application. They can be classified as follows:

- Direct dependencies - used explicitly by the application (e.g.: a JavaScript library imported by the developer).
- Transitive dependencies - automatically included through other libraries, without the developer being aware of their presence.
- Runtime dependencies - used in application execution (e.g. scripts loaded from a CDN - Content Delivery Networks).
- Build-time dependencies - involved in the compilation, testing or deployment process.
- The contemporary software ecosystem is also characterized by a series of structural features that amplify both the efficiency and vulnerability of modern applications:
- Recursive dependencies - a medium-sized project can include hundreds or even thousands of external packages, a phenomenon that exponentially increases the attack surface.
- Hierarchical complexity - the structure of dependency trees is often nested and opaque, making it difficult to map and manage them effectively.
- High dynamics of changes - frequent updates of external components require continuous monitoring, auditing and integrity validation processes.

These features, while contributing to the acceleration of innovation and flexibility of development, at the same time introduce an increased level of uncertainty and risk from a cybersecurity perspective.

In the JavaScript ecosystem, for example, an average project can include hundreds or even thousands of indirect dependencies, making it nearly impossible to manually verify the integrity of all components. Thus, poor dependency management becomes a structural vulnerability of the web ecosystem.

1.3. Vulnerabilities in software supply chain (software supply chain)

Software supply chain chain) represents the set of processes, resources and entities involved in the development and distribution of code. Vulnerabilities can occur at several levels: [14], [15], [16]

- **Compromising source code -** by injecting backdoors into popular libraries (e.g. attacks on npm packages).
- **Infrastructure compromise** - by hijacking a distribution service (e.g.: the Polyfill.io incident, where domain compromise led to the delivery of malicious scripts).
- **Design vulnerabilities** - lack of integrity checking mechanisms (Subresource Integrity, Content Security Policy).
- **Poor project ownership management** - changing or misappropriating an open- source domain/package (e.g. XZ Utils).

These vulnerabilities demonstrate that the security of modern applications depends not only on the organization's own code, but also on the entire ecosystem of integrated external components.

The diagram in the following image (Fig. 1) complements the above analysis, illustrating the typical flow of the software supply chain, from developers and maintainers (***developers who manage software projects, administrators, operators, etc.***) to the final application. The diagram highlights (Fig. 1) the critical points where vulnerabilities can be introduced and through which an attack can be propagated on a large scale.

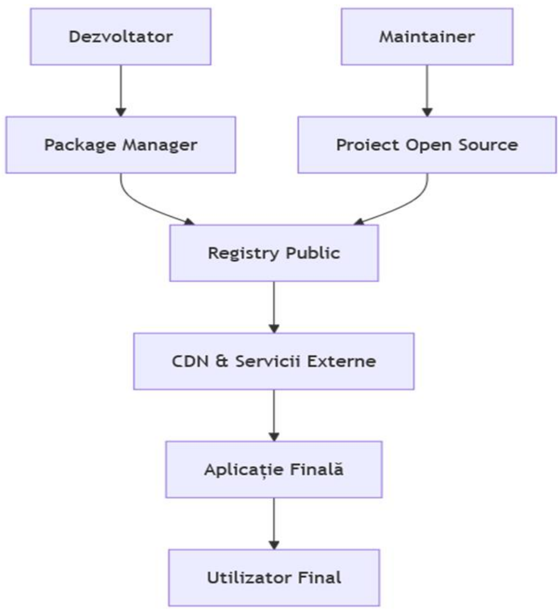


Fig. 1. Software supply flow chain and potential vulnerabilities

1.4. Methodology used (OSINT, technical documentation, public sources)

The present analysis was carried out through a mixed methodology, combining OSINT investigations with technical analyses and corroboration from official sources [1]-[9], [13], [14], [15].

Main directions:

1. **OSINT (Open Source Intelligence):** The collection, indexing, and correlation of information from public sources - technical reports and posts (e.g. Sansec, Cloudflare), press releases, technical articles, and discussions on social platforms (X/ Twitter, Reddit, Mastodon). This activity aims to identify indicators of compromise (IOCs), timelines of events, and community reactions [1], [3]-[9], [13].
2. **Technical documentation:** static and dynamic analysis of JavaScript code provided by Polyfill.io associated servers (download, search for redirect patterns), DNS/WHOIS investigations and examination of the described technical signatures [1], [3]-[5], [7], [8], [12].

3. This component includes **testing in isolated environments** (VMs) to confirm malicious behavior without exposing production systems.
4. **and official sources:** corroborating information with press releases and actions of institutional actors (e.g.: announcements by registry/domain operators, notifications by cloud /CDN providers, warnings issued by search engines or advertising platforms). This source allows for the assessment of the impact on a macro scale (market, reputation, provider involvement) [8], [9], [10], [11], [13], [15].

By combining the three dimensions, a comprehensive perspective is obtained, covering everything from fine technical details to the operational and economic consequences of the incident.

2. Technical analysis of the polyfill.io incident

Polyfill.io was an open- source **service** that provided JavaScript scripts to ensure compatibility of web applications with older browsers. By using this service, developers could automatically load only the functionality needed for a particular browser, optimizing performance and reducing maintenance effort [1]-[5].

The role of Polyfill.io in the web ecosystem

- **Universal compatibility** - allows modern websites to run on older browsers, expanding the audience.
- **Reducing developer effort** - instead of testing every feature on all browser versions, developers included a simple <script src ="https://polyfill.io/v3/polyfill.min.js">.
- **Standardization** - Polyfill.io has become a central point, adopted by **millions of sites** (government, commercial, educational), being perceived as an "infrastructure" service of the web.
- **Systemic risk** - it is precisely this centrality that created a critical dependency: compromising the polyfill.io domain meant instantly affecting a huge number of sites, without any action on their part.

Attack vectors and operating methods

The **incident did not directly compromise the original source code** of the *Polyfill project,* the project code (open-source, published on GitHub) remained untouched. The incident occurred after the polyfill.io **domain and associated CDN infrastructure** were taken over by a malicious actor (by purchasing the domain after it expired) who compromised them by *injecting malicious code* [1]-[5], [7], [8], [12].

Thus, the incident is classified as an **attack on the distribution infrastructure in the software supply chain**, not as an attack on the open- source code itself.

Once the attacker controlled **the CDN service**, they could:

- **Modify the delivered content** - instead of just providing legitimate polyfills, inject malicious JavaScript.
- **Personalize delivery** - malicious scripts were served selectively, depending on browser or region (conditional delivery), which made them harder to detect.
- **It exploited implicit trust** - millions of sites directly included <script src="https://polyfill.io/...">, without integrity checking.

3. Case study

Simulating the introduction of **the Array.prototype.findLast() function** in browsers: Chrome, Firefox, Opera and Internet Explorer and the polyfill risk, using **OSINT tools for analysis**.

To assess the behavior and risks associated with introducing certain functions JavaScript through *polyfill* mechanisms, a local testing platform was created in the form of a web page. This platform allowed verification of the existence and functionality of the Array.prototype.includes () method through *feature* techniques *detection*, as well as comparing the behavior when the function is provided natively by the browser engine versus when it is provided by an external polyfill.

The simulation consisted of two controlled scenarios:

- running the page without any additional script, to observe the native behavior of the browser.
- Deliberately loading a polyfill that implements or exposes the function globally, thus simulating the situation where a third-party script - potentially compromised - is present on the page.

The results were collected through screenshots, and the analysis focused on identifying the attack surface introduced by global polyfills.

This experimental architecture provides a reproducible and reliable basis for discussing the implications SAFETY supply-chain in the context of front- end dependencies.

Array.prototype.findLast () function was introduced in:

- Firefox version 143 (2025).
- Opera version 122 (2025).
- Not supported in Internet Explorer.
- Not supported in Chrome version 96 (2022).

In these browsers, the Array.prototype.findLast () function was identified using the query presented in Fig. 2:

```
<!doctype html>
<html>
<head><meta charset="utf-8"><title>Check findLast</title></head>
<body>
  <h2>Check Array.prototype.findLast()</h2>
  <pre id="out"></pre>
  <script>
    const out = document.getElementById('out');
    if ('findLast' in Array.prototype) {
      out.textContent = 'findLast disponibil nativ';
      console.log('findLast disponibil nativ');
    } else {
      out.textContent = 'findLast NU este disponibil — folosește shim/ponyfill';
      console.log('findLast NU este disponibil — folosește shim/ponyfill');
    }
  </script>
</body>
</html>
```

Fig. 2. The Array.prototype.findLast ()

The results are shown in the following images (Fig. 3- Fig. 6):

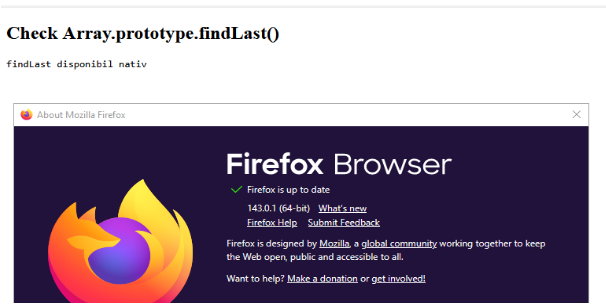


Fig. 3. Checking the existence of the function in the browser Firefox

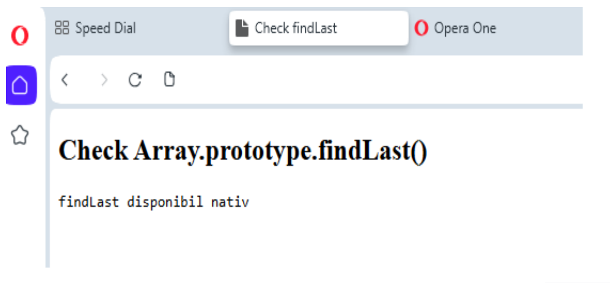


Fig. 4. Checking the existence of the function in the Opera browser

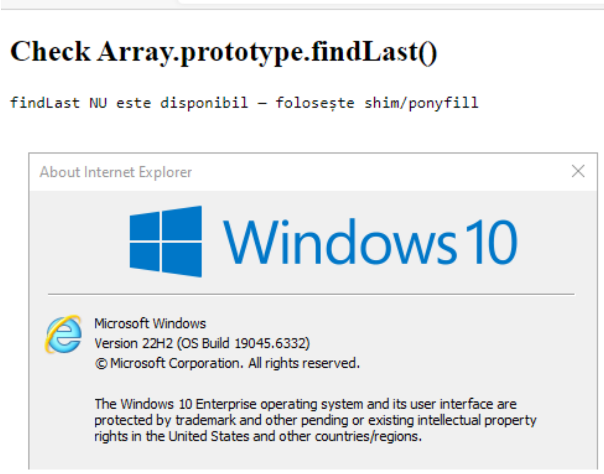


Fig. 5. Checking the existence of the function in the Internet Explorer browser

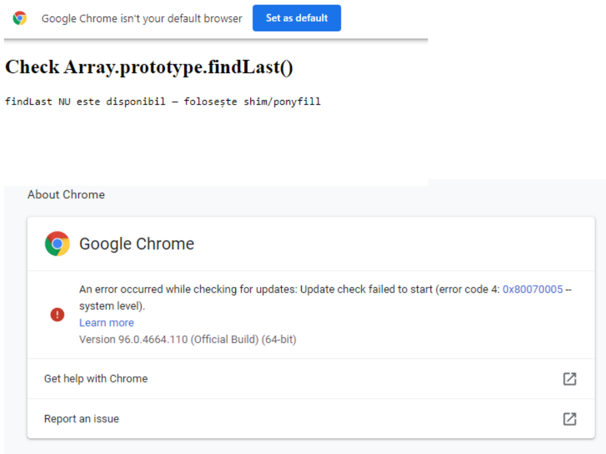


Fig. 6. Checking the existence of the function in the browser Chrome version 96

Let's assume that the JavaScript function array.prototype.includes() is standard. Newer browsers (Chrome 87+, Firefox 43+) implement it, but older browsers (Internet Explorer) do not. A polyfill for this function would check if the function already exists in the browser. If it doesn't, it *creates it* using backward-compatible JavaScript code (Fig. 7 - Fig. 10).

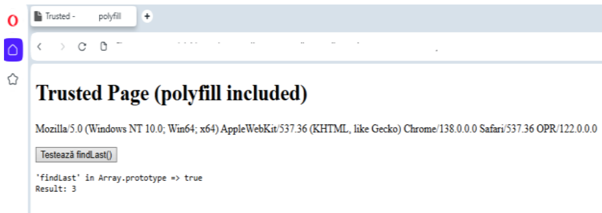


Fig. 7. Firefox/143.0

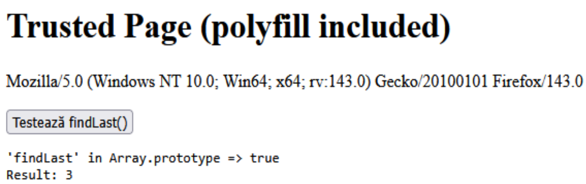


Fig. 8. Chrome/138.0.0.0

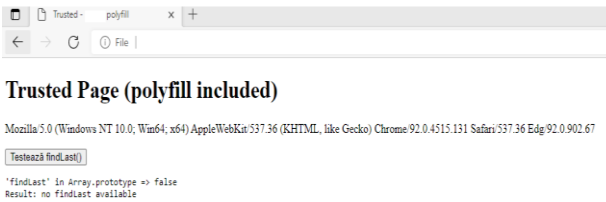


Fig. 9. Chrome/92.0.4515.131

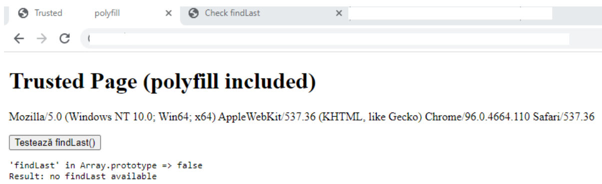


Fig. 10. Chrome/96.0.4664.110

Simulating a page that loads the polyfill (simulates a page that could abuse it if the polyfill is global). The code run is presented in Fig. 11.

After running the code in the browser Chrome v.96 (*no* native implementation); result is false.

The polyfill is not installed globally - the "safe" version is running which does not expose the method globally (deliberate intention for security).

ponyfill / shim was used which **does not modify** Array.prototype (and then findLast does not appear globally, it remains available only through the helper).

The result " no global findLast " indicates that a malicious page cannot call arr.findLast (...) directly (in this session), so this concrete attack surface does not exist at the time of capture.

In Fig. 12, the attacker *does not* have access to the global API.

```html
<!doctype html>
<html lang="ro">
<head>
  <meta charset="utf-8">
  <title>Attacker - demo polyfill</title>
  <meta name="viewport" content="width=device-width,initial-scale=1">
  <!-- Attacker includes the same script (simulating a compromised CDN) -->
  <script src="findLast-polyfill.es5.js"></script>
</head>
<body>
  <h1>Attacker Page (polyfill included)</h1>
  <p id="status"></p>
  <button onclick="tryAbuse()">Try abuse (read global)</button>
  <pre id="out"></pre>
  <script>
    function tryAbuse(){
      const out = document.getElementById('out');
      out.textContent = '';
      try {
        out.textContent += "'findLast' in Array.prototype => " + ('findLast' in Array.prototype) + "\n";
        if ('findLast' in Array.prototype) {
          var arr = ['a','b','c'];
          var res = arr.findLast(x => x === 'b');
          out.textContent += 'Attacker read result: ' + String(res) + '\n';
          console.log('Attacker read:', res);
        } else {
          out.textContent += 'no global findLast';
        }
      } catch(e){
        out.textContent += 'Error: ' + e.message;
      }
    }
    document.getElementById('status').textContent = navigator.userAgent;
  </script>
</body>
</html>
```

Fig. 11. Simulating a page that loads the polyfill

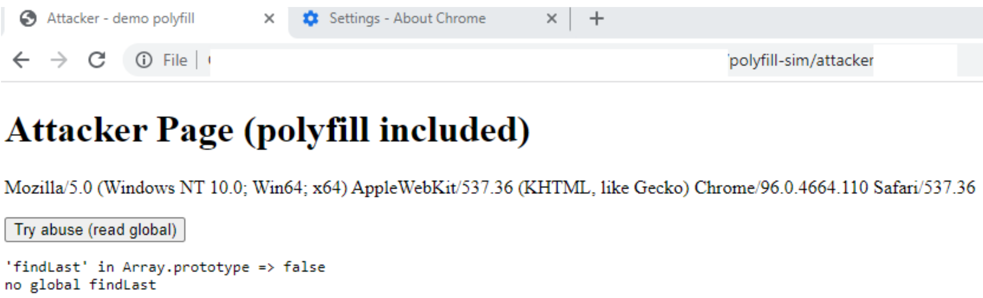


Fig. 12. The attacker *does not* have access to the global API

After running the code in the browser Firefox (v ~143) notices that *findLast* is present in Array.prototype (value is **true**) (Fig. 13).

This value can be the result of:

- **The browser 's native implementation** (in your case Firefox 143, findLast is natively supported from versions much older than 143).
- **Polyfill that installed the method globally** (if you loaded an external script that had modulated Array.prototype).

Fig. 13. The running the code in the browser Firefox (v ~143)

Since the user -agent indicates Firefox v143 (updated browser), most likely findLast is **native**, but we cannot rule out the fact that the page also loaded a legitimate polyfill (which would also indicate the value true when calling *findLast*).

OSINT elements used in the study

In this study, the analysis of polyfill adoption and exposure was based on a structured set of *Open Source Intelligence (OSINT) techniques* and tools. The goal was to obtain verifiable and reproducible evidence of the use of polyfill functions. JavaScript (in particular Array.prototype.findLast) and polyfill distribution mechanisms, without resorting to intrusive or access-violating actions [1]-[9], [13]-[15].

The analysis capitalized on a set of complementary sources:

- repositories (e.g. GitHub).
- Package registry (e.g.: npm) - consulted to detect packages that provide shim / ponyfill and to understand the distribution mode (dedicated packages vs. CDN inclusions).
- archives and indexes (e.g. Common Crawl, Wayback Machine) - used for quantitative estimates: frequency the appearance of references to certain CDNs in historical or current web pages.
- Search engine - for identifying mentions in articles, issues and technical discussions, useful in contextualizing decisions.

Certificate transparency / crt.sh - to check related events related to domains hosting polyfills or CDNs (only for passive mapping of public domains).

4. Conclusions

The research exclusively targeted public data and open resources; any private or restricted information accidentally identified was ignored and not collected or stored.

Results from the analysis were written to support risk management, not to facilitate exploitation. OSINT tools, used within an ethical framework and reproducible, allow the identification and assessment of polyfill exposure and distribution practices in the web ecosystem.

The described methodology provides a solid basis for informative estimates, useful both for research/investigation and for risk management plans at the organizational level [14], [15], [16].

References

[1]. Sansec Forensics Team, "Polyfill supply chain attack hits 100K+ sites," Sansec, technical analysis and timeline of the compromise affecting the polyfill.io domain and the injection of malicious code via CDN scripts, 2024.

[2]. A. Sharma, "Polyfill.io supply chain attack hits 100,000 websites - all you need to know," Sonatype, blog post summarizing impact and initial recommendations for organizations, 2024.

[3]. J. Graham-Cumming and Cloudflare Team, "Automatically replacing polyfill.io links with Cloudflare's mirror for a safer Internet," The Cloudflare Blog, discusses mirroring/URL rewrite and practical mitigations, 2024.

[4]. Akamai Security Blog, "Examining the Polyfill attack from Akamai's point of view," Akamai, technical analysis and CDN/security provider perspective, 2024.

[5]. Snyk Research, "Polyfill supply chain attack - analysis & mitigation guidance," Snyk, contextualizes the incident in the JavaScript ecosystem and proposes fixes and scanning approaches, 2024.

[6]. S. Sarva, “Polyfill.io supply chain attack: what you need to know,” Qualys, technical note with detection and scanning recommendations for organizations, 2024.

[7]. Censys Threat Research, “Polyfill.io supply chain attack - digging back the web of compromised domains,” Censys, investigation mapping domains and hosts involved, 2024.

[8]. FortiGuard Labs, “Polyfill.io Supply Chain Attack - Threat Signal Report,” Fortinet, summary of indicators of compromise (IoCs) and recommendations, 2024.

[9]. The Hacker News, “Polyfill.io attack impacts 380,000+ hosts,” The Hacker News, journalistic coverage of scope and follow-ups, 2024.

[10]. NVD, “CVE entries related to Polyfill.io usage (e.g. CVE-2024-38526),” NIST National Vulnerability Database, technical CVE details and references to affected packages, 2024.

[11]. Cloudflare, “Automated rewrites & mirrors; guidance for customers to remove polyfill.io,” The Cloudflare Blog, describes rapid large-scale mitigation steps, 2024.

[12]. FOSSA, “Polyfill supply chain attack: details and fixes,” FOSSA Blog, explains attack mechanics and remediation steps for development teams, 2024.

[13]. Checkmarx, Invicti, Arctic Wolf, Kaspersky, “Polyfill.io supply chain attack - vendor advisories and practical recommendations,” multiple vendor security advisories, 2024.

[14]. OWASP, “Software Supply Chain Security Cheat Sheet,” OWASP Cheat Sheet Series, practical recommendations for assessing and mitigating software supply chain risks (SCA, SRI, self-hosting, SBOM, etc.), 2023.

[15]. NIST, “Software Supply Chain Security Guidance (EO 14028),” NIST guidance documents on policies and standards for software supply chain security, 2021-2024.

[16]. “On the prevalence of software supply chain attacks,” and related academic studies on supply chain attack taxonomy and investigation frameworks, 2020-2024.

Enhancing 5G Infrastructure to Withstand Emerging Digital Threats

Andreea BENCHEA
Faculty of Electronics, Telecommunications and Information Technology,
National University of Science and Technology POLITEHNICA Bucharest, Romania
andreea.benchea@stud.etti.upb.ro

Abstract

The advent and rapid expansion of 5G technology brings substantial advancements in communication capabilities, characterized by ultra-low latency, enhanced bandwidth, and massive device connectivity. However, this technological evolution simultaneously exposes critical infrastructure to a broad spectrum of sophisticated and evolving digital threats. This paper addresses the security challenges inherent to 5G networks and proposes a set of advanced, intelligent solutions tailored to their architectural complexity. The proposed measures include Zero Trust Architecture (ZTA), artificial intelligence-based behavioral analytics, federated learning, blockchain-enabled device authentication, and secure orchestration of network slicing. These methodologies offer a scalable, proactive, and privacy-conscious security framework capable of ensuring operational resilience and data integrity. The objective of this work is to emphasize the necessity of adopting adaptive and future-ready defense mechanisms to safeguard the robustness and reliability of 5G infrastructures against emerging cyber threats.

Index terms: 5G security, Artificial Intelligence, cyber threats, network slicing, Zero Trust Architecture

1. Introduction

Fifth-generation mobile networks represent a pivotal step in the evolution of global telecommunications infrastructure, offering capabilities that significantly surpass those of previous generations. With enhanced speed, lower latency, and the ability to support a massive number of connected devices simultaneously, 5G is poised to facilitate the growth of transformative technologies, including autonomous systems, remote surgery, real-time analytics, and large-scale smart city deployments [1].

Beyond its transformative potential, 5G introduces a paradigm shift in the architecture of communication systems. Unlike its predecessors, 5G is built upon a software-centric framework that incorporates virtualization, cloud-native infrastructure, and edge computing. These characteristics enable agility and scalability but also open the door to new classes of attacks that target dynamic network functions, orchestration layers, and APIs. This new complexity demands continuous visibility, real-time threat detection, and automated defense mechanisms.

In parallel, the vast expansion of connected devices, from personal wearables to industrial IoT endpoints, generates an explosion of data, much of it sensitive or mission-critical. In this context, ensuring data confidentiality, integrity, and availability becomes paramount. The convergence of AI, automation, and 5G network services requires security solutions that are not only robust but also context-aware, adaptive, and aligned with privacy-preserving principles. The challenge is no longer only about preventing breaches but about architecting inherently resilient systems.

However, the widespread implementation of 5G infrastructure also brings about new and complex cybersecurity concerns. As the architecture becomes more open and distributed, new attack surfaces emerge, affecting everything from core network components to edge devices and user endpoints. Moreover, the sheer scale of interconnectivity envisioned by 5G increases the likelihood of vulnerabilities being exploited and threats propagating rapidly across systems [2].

To address these risks, the security of 5G networks must be treated as a foundational design principle rather than a peripheral consideration. Traditional, static security approaches are no longer sufficient in a hyperconnected, real-time environment. Instead, proactive and adaptive security mechanisms are required to ensure resilience, privacy, and trust in next-generation mobile networks.

2. Threat Landscape in 5G Networks

The architecture and operational complexity of 5G introduce multiple vectors for sophisticated digital threats. Below are key categories of risks identified in current 5G deployments:

- **Virtualized Network Functions (VNFs)**: The use of VNFs-software-defined equivalents of traditional network functions-poses significant challenges. VNFs may be exploited through techniques such as malware injection, insecure APIs, configuration errors, or privilege escalation attacks. They also introduce new dependencies on software supply chains, increasing systemic risk [3].
- **Edge Computing Nodes**: Edge computing decentralizes processing by relocating computational tasks closer to end users and devices. While this improves performance and reduces latency, it creates localized targets for attackers. Edge nodes often operate in less secure environments and are susceptible to data interception, spoofing, and unauthorized access [4].
- **Supply Chain Vulnerabilities**: The involvement of multiple vendors in building and maintaining 5G networks expands the attack surface via potentially insecure hardware, firmware backdoors, or compromised development processes. These vulnerabilities can be exploited to gain long-term, undetected access to core infrastructure [5].
- **Network Slicing Attacks**: 5G allows operators to create isolated logical networks (slices) tailored for different services. However, improper slice isolation or insecure orchestration may lead to data leakage, denial of service (DoS), or privilege escalation across slices [6].
- **IoT Exploitation**: 5G is expected to connect tens of billions of IoT devices. Many of these devices lack strong security configurations and are often deployed in high-risk environments. Exploiting insecure IoT endpoints, attackers can launch botnet-based DDoS attacks or pivot into critical infrastructure systems [7].

The scale, heterogeneity, and real-time nature of 5G amplify these threats, necessitating intelligent, context-aware, and proactive defense mechanisms that go beyond static firewalls and manual monitoring.

3. Advanced Security Solutions for 5G

3.1. Zero Trust Architecture (ZTA)

Zero Trust Architecture is based on the principle of "never trust, always verify." In the context of 5G, ZTA eliminates implicit trust between devices, services, and users by enforcing strict identity verification, real-time monitoring, and access control regardless of location. It is particularly relevant for microservices, virtualized network functions (VNFs), and mobile edge environments [2][3].

For example, a telecom operator could implement ZTA by requiring all components-whether in the core network or at the edge-to authenticate through a centralized identity and policy enforcement system. If a rogue base station attempts to connect to the network, the access request will be evaluated against device identity, behavior history, geolocation, and role-based permissions. If any anomaly is detected (e.g., unusual login time, inconsistent firmware signature, or unregistered IP range), access is denied, and a security alert is triggered. This prevents lateral movement by potentially compromised nodes and ensures that access is strictly limited to verified, authorized actors.

3.2. AI-Based Behavioral Analytics (UEBA)

User and Entity Behavior Analytics (UEBA) applies machine learning to establish baselines of normal behavior and detect deviations that may indicate insider threats, advanced persistent threats (APTs), or misconfigurations. In 5G networks, UEBA can be deployed across slices and virtualized components to monitor network traffic patterns and user behavior, identifying threats in near real-time and automating response actions [2][5].

As an example, consider a smart grid operator managing power distribution across multiple cities via a dedicated 5G slice. Under typical conditions, devices follow consistent communication schedules and interact with known control centers. If a UEBA system detects that one sensor node suddenly initiates large volumes of outbound connections to foreign IP addresses during off-peak hours, it flags the event as anomalous. The platform then isolates the device, alerts administrators, and initiates a forensic analysis to determine whether the device has been compromised or misconfigured.

3.3. Federated Learning (FL)

Federated learning enables AI models to be trained across multiple edge devices without the need to share raw data with a central server, thus preserving user privacy and reducing communication overhead. In 5G, where edge computing is prominent, FL allows distributed threat detection and adaptive security intelligence while keeping sensitive data localized [4].

A practical example might involve multiple hospitals using connected medical devices such as infusion pumps, monitors, or diagnostic equipment. Rather than uploading patient data to the cloud, which could expose sensitive health information, each device uses its local data to improve a machine learning model that detects device malfunctions or security anomalies. The model updates (not the data) are sent to a central aggregator, which merges them into a global model and sends it back to all devices. This allows the network to learn from localized incidents (e.g., attempted ransomware injections on a single device) and preempt similar attacks elsewhere, without breaching patient confidentiality.

3.4. Blockchain for Device Authentication and Integrity

Blockchain technology provides a tamper-proof mechanism for tracking authentication events and securing device identities. In 5G infrastructures, blockchain can be applied to validate software updates, authenticate connected devices, and maintain secure records of configuration changes. This enhances transparency and accountability while mitigating risks associated with spoofing and unauthorized access [5][7].

For instance, a mobile operator may maintain a private blockchain ledger where every network device, from base stations to IoT sensors, must register and verify its identity. When a firmware update is issued, the cryptographic hash of the update package is recorded on the blockchain. Devices downloading the update validate it against the hash. If a malicious actor attempts to introduce a tampered update (e.g., with embedded spyware), the hash mismatch alerts the blockchain node, and the update is rejected. This decentralized trust mechanism ensures end-to-end integrity across highly distributed 5G ecosystems.

3.5. Secure Network Slicing Orchestration

Network slicing enables operators to allocate isolated virtual networks to support diverse applications. However, without secure orchestration, vulnerabilities in one slice could compromise others. AI-powered orchestration platforms can dynamically enforce access controls, monitor performance, and isolate threats within individual slices. According to 3GPP specifications, strong authentication, integrity protection, and isolation are essential to prevent cross-slice exploits [6].

For example, consider a telecom provider offering separate network slices to autonomous vehicle fleets, public safety services, and streaming media. If the orchestration system detects that traffic in the vehicle fleet slice suddenly spikes with patterns resembling a DDoS attack, it automatically reroutes critical telemetry to backup slices, applies stricter firewall rules, and temporarily isolates the impacted slice. Meanwhile, other services continue operating without degradation. This segmentation ensures that malicious activity in one slice does not cascade into other mission-critical services, maintaining overall network reliability and trust.

4. Regulatory, Ethical and Operational Considerations

Securing 5G infrastructure is not solely a technical challenge, it requires alignment with legal frameworks, ethical guidelines, and operational strategies that span national and international domains. At the regulatory level, one of the key instruments in the European Union is the EU Toolbox for 5G Security, which outlines a coordinated set of risk mitigation measures for member states, addressing aspects such as supply chain control, access restriction, and secure software lifecycle management [5]. Complementarily, the NIS2 Directive expands the cybersecurity obligations for essential and important entities, including telecom operators, by imposing stricter incident reporting, risk assessment, and supply chain security requirements [8].

Globally, frameworks from organizations such as ITU, 3GPP, and ENISA advocate for secure-by-design principles, network resilience, and proactive threat management. These align with the growing emphasis on adopting Zero Trust principles and multi-layered defense architectures.

From an ethical standpoint, the implementation of AI-based systems within 5G security raises important concerns. Key issues include:

- Data privacy: Algorithms such as UEBA and federated learning require access to potentially sensitive metadata. Compliance with GDPR and similar data protection regulations is essential.
- Algorithmic bias: AI models trained on unbalanced or incomplete datasets may lead to biased threat assessments or unjustified access denials.
- Transparency and accountability: The use of opaque decision-making processes must be countered by explainable AI (XAI) methods and clear human oversight protocols.

Operationally, integrating the aforementioned solutions into real-world 5G networks demands a significant upgrade in infrastructure, processes, and human competencies. Network operators must:

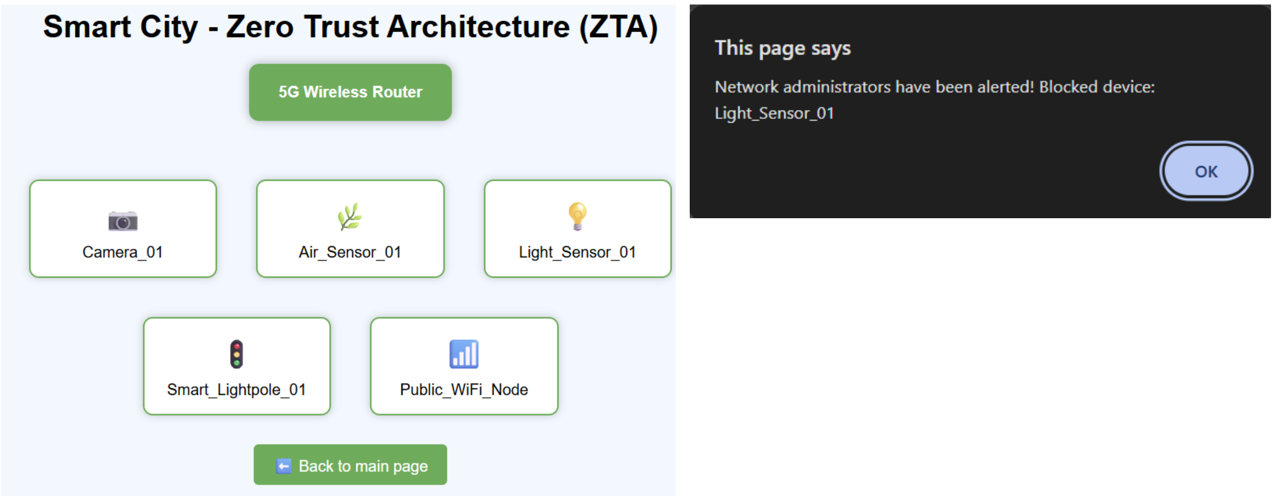
- Invest in automation and orchestration platforms capable of responding to threats in real-time.
- Establish inter-industry and public-private collaboration frameworks to share threat intelligence and mitigation strategies.
- Train personnel in AI operations (AIOps), blockchain governance, and Zero Trust deployment models.

A robust 5G security posture thus requires a convergence of technical excellence, ethical diligence, and regulatory compliance.

5. Case study - Applying 5G security solutions in a smart city scenario

This study presents the practical deployment of 5G security strategies within a simulated smart city infrastructure. The developed model connects critical systems including public transportation, traffic signaling, video surveillance, environmental monitoring sensors, and emergency response units via a 5G network.

Fig. 1. Overview of 5G Security Strategies Implemented in the Smart City Architecture



(a) (b)

Fig. 2. (a) Device authentication interface; (b) Security alert triggered after device blocking

The initial stage involved the implementation of a Zero Trust Architecture (ZTA) across all connected assets. Each network component, such as surveillance cameras and connected streetlights, is required to verify its identity prior to accessing the infrastructure. Unauthorized devices are systematically denied access, and lateral communication between assets is permitted solely under predefined and encrypted policies, thereby significantly mitigating internal threat propagation. Within the associated simulation, network access is conditioned by password verification. Devices failing authentication after three attempts are visually blocked, and a network alert is triggered, emphasizing proactive internal security measures.

Subsequently, AI-based behavioral analytics were integrated both at the network core and edge levels to establish normative activity baselines. These systems continuously monitor real-time device interactions, detecting deviations from expected behavior. For example, if a waste sensor transmits data during unusual hours or towards unauthorized endpoints, an alert is generated and the device may be isolated for further investigation. In the simulation, real-time charts are displayed for smart sensors. Data points exceeding defined thresholds are automatically highlighted in red, facilitating the immediate detection of anomalies.

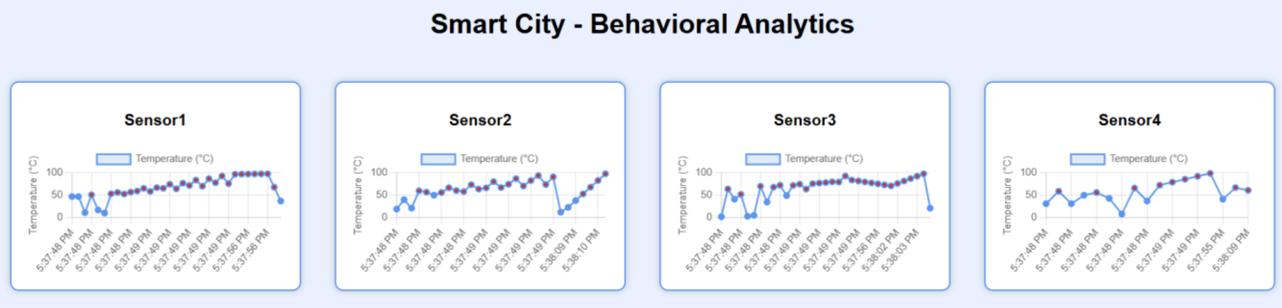
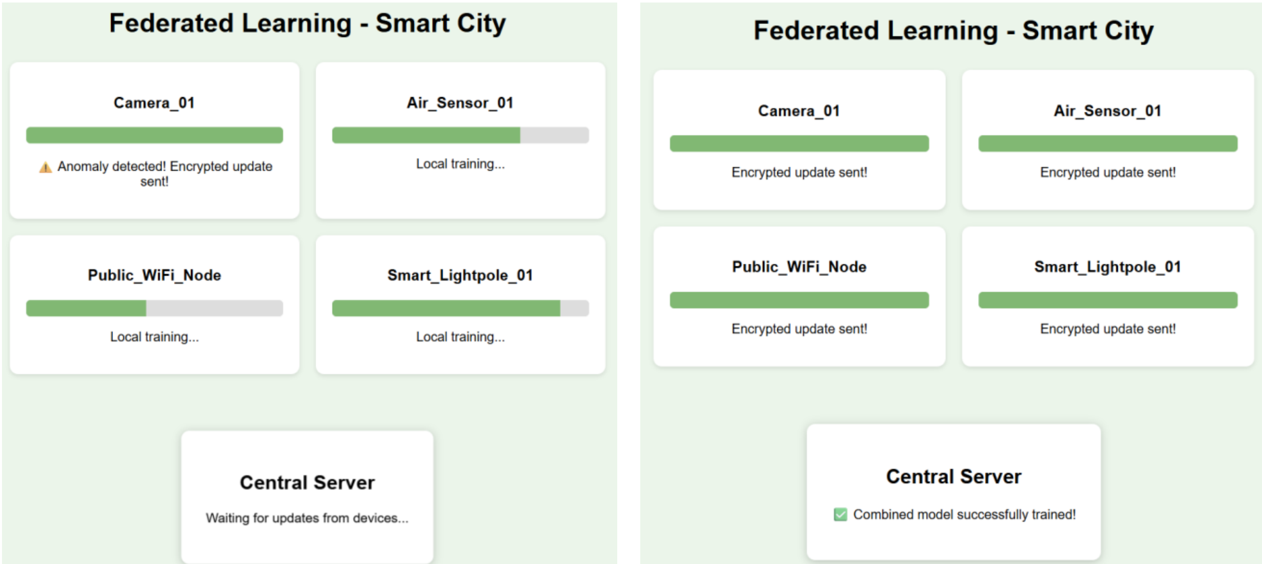


Fig. 3. Real-Time Behavioral Analytics Monitoring for Smart City Sensors

In order to enhance localized threat intelligence without compromising user privacy, Federated Learning methodologies were employed. Smart meters and environmental monitoring units train machine learning models locally and transmit solely encrypted model updates to a central aggregation node, thus maintaining data sovereignty. In the corresponding simulation, devices illustrate the local model training process and securely transmit encrypted updates. This approach ensures private data remains on the device while contributing to an improved global model.



(a) (b)

Fig. 4. (a) Local training phase with anomaly detection; (b) Successful encrypted updates and combined model training at the central server

To preserve the authenticity and integrity of devices and software updates, a blockchain-based authentication system was integrated. This immutable ledger maintains records of device registrations, configuration changes, and software update histories. Prior to the application of any update, devices verify the corresponding hash against the blockchain to prevent tampering or unauthorized modifications. Within the simulation, update hashes are validated against existing blockchain records. Updates matching the recorded hash are successfully applied, while those failing verification are promptly rejected.

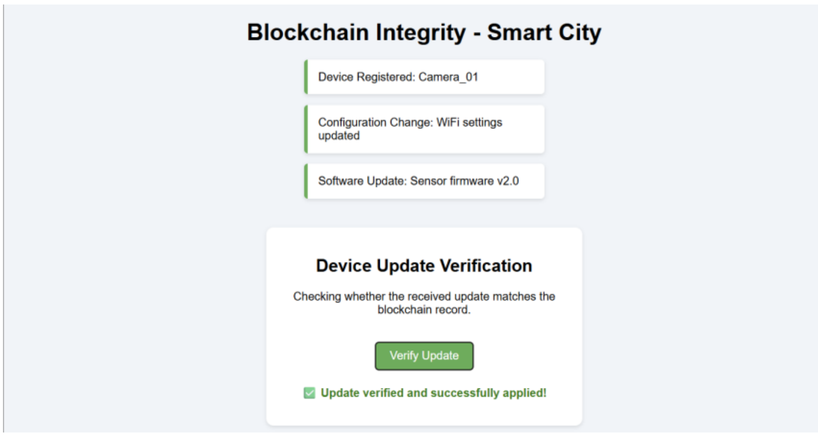


Fig. 5. Blockchain-Based Device Update Verification

In order to optimize service delivery and security, secure network slicing was employed to segment network services according to function, specifically public connectivity, emergency response, utilities, and traffic management. Each slice operates independently under the supervision of AI-based orchestration platforms, which manage resource allocation and enforce security policies. In the event of a cyberattack affecting one slice (such as a denial-of-service attack on public Wi-Fi), the remaining slices continue to operate normally, ensuring service continuity without cross-contamination. In the simulation, random attack events can be simulated, visually isolating the affected slice while maintaining normal operations across unaffected slices, thus demonstrating the resilience of a segmented network architecture.

Fig. 6. Simulation of Attack Detection and Isolation in Smart City Network Slicing

6. Conclusions and future directions

The evolution of 5G networks marks a transformative milestone in modern communications, enabling capabilities such as ultra-reliable low-latency communication, massive machine-type communication, and enhanced mobile broadband. However, with these advancements come unprecedented cybersecurity risks due to the network's decentralized architecture, software-defined components, and hyper-connectivity.

This paper has examined the unique threat vectors introduced by 5G technologies and outlined a comprehensive portfolio of security mechanisms - including Zero Trust Architecture, behavioral analytics, federated learning, blockchain-based integrity verification, and secure network slicing - that collectively support a robust, adaptive, and proactive security posture.

As technology adoption accelerates, it is imperative to continue evolving these frameworks. Future efforts must prioritize:

- Developing explainable and auditable AI models that ensure transparency and traceability in automated threat detection.
- Standardizing and regulating blockchain implementations tailored to telecom requirements, ensuring they align with global interoperability and data governance policies.
- Securing the AI development lifecycle, particularly against poisoning and adversarial inputs at the edge.
- Fostering international cooperation, including harmonized cybersecurity standards, collaborative incident response protocols, and cross-border threat intelligence sharing.

The long-term security of 5G infrastructures necessitates sustained investment in specialized education, continuous workforce development, and the dynamic adaptation of regulatory frameworks to technological advancements. Safeguarding these networks constitutes an ongoing endeavor, requiring unwavering technological vigilance, operational resilience, and a firm commitment to ethical governance across all layers of the digital ecosystem.

References

[1]. International Telecommunication Union (ITU), *IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond*, Recommendation ITU-R M.2083-0, Sep. 2015. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I%21%21PDF-E.pdf.

[2]. European Union Agency for Cybersecurity (ENISA), *Threat Landscape for 5G Networks*, Nov. 2020. [Online]. Available: https://www.enisa.europa.eu/sites/default/files/publications/ENISA%20threat%20landscape%20for%205G%20Networks.pdf.

[3]. 5GPPP, Security *Landscape - Phase 1 White Paper*, Jun. 2017. [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP_White-Paper_Phase-1-Security-Landscape_June-2017.pdf.

[4]. European Telecommunications Standards Institute (ETSI), *Multi-access Edge Computing (MEC); Framework and Reference Architecture*, ETSI GS MEC 003 V2.1.1, Jan. 2019. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/02.01.01_60/gs_MEC003v020101p.pdf.

[5]. European Commission, *Cybersecurity of 5G networks - EU Toolbox of risk mitigating measures*, Jan. 2020. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/cybersecurity-5g-networks-eu-toolbox-risk-mitigating-measures.

[6]. 3GPP, *TR 33.891 - Study on security aspects of network slicing in 5G networks*, Release 15, Dec. 2018. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3358.

[7]. Symantec, *Internet Security Threat Report*, Vol. 24, Feb. 2019. [Online]. Available: https://docs.broadcom.com/doc/istr-24-2019-en.

[8]. European Union, *Directive (EU) 2022/2555 (NIS2)*, Official Journal of the European Union, Dec. 2022. [Online]. Available: https://eur-lex.europa.eu/legal-content/RO/TXT/?uri=CELEX%3A32022L2555.

Law Enforcement Cooperation in the Prevention and Countering of Disinformation

Marius-Andrei OROȘANU
Police Academy „A.I. Cuza" Bucharest, Romania
andrei.orosanu@academiadepolitie.ro

Abstract

The growing impact of disinformation on public trust, democratic processes, and national security has made the prevention and countering of disinformation a strategic priority for law enforcement agencies. This paper explores the role of police cooperation -both national and international - in identifying, preventing, and responding to disinformation campaigns, particularly those amplified through digital platforms and social media. Law enforcement authorities, in collaboration with cybersecurity units, must adapt their operational frameworks to detect and address coordinated information manipulation. Romania, as a member of the European Union and signatory to multiple international agreements, actively participates in joint operations, data sharing, and institutional efforts through bodies such as Europol, Eurojust, and the European Centre of Excellence for Countering Hybrid Threats. Additionally, the involvement of civil society, media organizations, and private tech companies is essential for building resilience against disinformation. The paper argues that successful prevention requires a multidimensional approach: legal harmonization, technical capacity building, and the strengthening of cross-border cooperation mechanisms. In this context, Romania's growing institutional capability and engagement in EU-led initiatives underline its strategic role in the regional fight against disinformation.

Index terms: cybercrime, disinformation, European Union, hybrid attack, law enforcement cooperation

1. Introduction

In contemporary society, individuals are constantly exposed to a wide array of situations in which their responses are shaped by how they perceive what they see, hear, and read. Despite widespread access to education and digital technologies - many of which are specifically designed to support learning and self-directed knowledge acquisition - individuals are increasingly confronted with informational content that persuades them to accept presented narratives as inherently true.

This phenomenon has become particularly prominent in recent years, driven by several key factors: limited perceptual discernment, inadequate levels of formal education and self-education, an underdeveloped regulatory framework addressing information integrity, the complex and often manipulative nature of the media environment, and also the accelerated process and widespread dissemination of deceptive or misleading informational content.

These dynamics raise concerns regarding the public's ability to assess information, and they underscore the need for comprehensive strategies aimed at fostering media literacy, promoting institutional accountability, and strengthening individual resilience against disinformation.

In response to these challenges, national and international authorities have intensified their efforts to curb the spread of disinformation and promote critical thinking within the public sphere.

Through collaborative initiatives involving governmental institutions, educational systems, media organizations, and civil society, there is a growing emphasis on the importance of being accurately informed. These efforts aim not only to limit the reach of false or misleading content, but also to instill a culture of information verification, encouraging individuals to assess the credibility of sources before accepting or sharing information that may, directly or indirectly, impact their daily lives. The success of such initiatives depends largely on sustained public engagement and a shared understanding that being well-informed is both a personal responsibility and a societal necessity.

2. Definition and terminology

Disinformation, as a concept and term, is defined by competent institutional authorities as a harmful and growing phenomenon within society. It is understood as a form of strategic communication involving the deliberate dissemination of false or misleading information - whose inaccuracy can be objectively verified - with the purpose of obtaining economic gain or manipulating public opinion. The impact of such actions can lead to significant harm at the social, political, or institutional level, undermining democratic processes, public policymaking, and collective goods such as public health, security, or the environment. A piece of content may be classified as disinformation when the following essential criteria are cumulatively met:

- The information presented is false or misleading, and this nature can be objectively verified through reliable means of fact-checking.
- The intent behind its dissemination is either to secure economic gain or to deliberately manipulate public perception.
- The content possesses the potential to cause harm at a collective level, such as undermining democratic processes, compromising public safety, or eroding trust in state institutions.

The concept of disinformation, as defined at the European Union level, explicitly excludes certain forms of expression that, while potentially containing inaccuracies or provocative elements, do not fall within the scope of disinformation. These include:

- Satire and parody, recognized as legitimate forms of artistic expression and social commentary.
- Partisan commentary, provided it is clearly identifiable as such and does not involve deliberate deception.
- Editing or citation errors, which lack manipulative intent.
- Commercial advertising, which is regulated separately through specific legislation and self-regulatory codes within the advertising sector.

Although today’s technological landscape is markedly different, understanding the concept of “informational dysfunction” requires a historical perspective on how information has been manipulated - through propaganda, strategic disinformation, satire, or hoaxes - as a persistent feature of the communication ecology. The term refers to a systemic imbalance in the information environment, wherein the boundaries between verified content and fabricated narratives become increasingly blurred, thereby impairing the public’s ability to distinguish between genuine public-interest information and manipulative messaging.

Within this context, professional journalism plays a critical role in addressing such dysfunctions; however, it remains vulnerable, given that the development of journalistic standards is relatively recent when compared to the long-standing history of information manipulation [1].

The terms “fake news” and, more recently, “fake media” have undergone significant semantic distortion, being frequently misused to label information, institutions, or sources with which certain actors - often political - disagree, rather than referring to content that is factually incorrect, fabricated, or deliberately misleading. This improper usage has led to a conceptual erosion of the term, undermining its analytical clarity and weakening the capacity to define and address harmful informational phenomena with precision. Data extracted from Google Trends indicate a marked increase in public interest for the term "fake news" beginning in the second half of 2016 - a period that coincides with its accelerated incorporation into global media and political discourse.

This analytical module aims, on one hand, to argue that the expression "fake news" is inadequate for capturing the complexity and scale of contemporary information pollution, and on the other hand, to underscore the problematic and potentially damaging consequences of its continued use in public discourse.

A key concern lies in the term's structural susceptibility to politicization and rhetorical manipulation. "Fake news" has increasingly been deployed as a discursive weapon - used to delegitimize media institutions, discredit professional journalism, and erode public trust in independent sources of information. In light of this trend, experts in public communication and information policy advocate for abandoning the term in favor of more precise and operationally valid concepts such as "disinformation" - the intentional dissemination of false information with manipulative intent - and "misinformation", which refers to the unintentional spread of inaccurate information absent deliberate intent to mislead [2].

3. The legal framework and its effectiveness in a rule-of-law state

In light of the accelerated development of digital technologies and the increased accessibility of informational content, including disinformation, EU Member States and their partner countries have recently established a more structured and legally coherent framework to address these challenges. Although earlier legal instruments - such as directives, regulations, and non-binding guidelines - existed, the current informational landscape demands centralized and systematic updates. Without such reform, societies remain vulnerable to high volumes of content that deviate from accepted standards of public discourse. As a result, intergovernmental legislative cooperation has been reinforced, and the existing legal architecture has undergone critical reassessment and targeted refinement to enhance both its efficiency and normative adequacy.

The present Regulation establishes a fully harmonised legal framework for intermediary services across the internal market, aiming to create a secure, transparent, and trustworthy online environment. It addresses the proliferation of illegal content and the broader societal risks posed by the dissemination of disinformation and other harmful materials, while ensuring the effective protection of fundamental rights as laid down in the Charter and promoting digital innovation. Consequently, Member States are precluded from introducing or upholding national provisions in areas covered by this Regulation, unless explicitly authorised to do so. Such unilateral measures would risk undermining the uniform and direct application of the harmonised rules governing intermediary service providers. However, this does not exclude the possibility for Member States to enforce other national laws - consistent with Union law, including Directive 2000/31/EC, particularly Article 3 - where those laws pursue legitimate public interest objectives distinct from those addressed by this Regulation.

The Regulation categorizes the risks related to disinformation into four principal groups. These encompass:

- The first category of risks involves the spread of illegal content such as child sexual abuse material, unlawful hate speech, and other criminal uses of online services. It also includes the trade of illegal goods like counterfeit products and illicit animal trafficking. These activities pose systemic risks, especially when amplified by widely influential accounts or mechanisms. Very large online platforms and search engines must assess these risks, regardless of their terms of service. This risk assessment does not affect the legal liability of users or website operators under applicable law.

- The second category concerns the impact of digital services on fundamental rights protected by the EU Charter, including human dignity, freedom of expression, privacy, non-discrimination, children's rights, and consumer protection. Risks can stem from algorithm design or abusive practices such as malicious takedown notices or silencing users. Providers must especially address risks to minors, including exposure to harmful or addictive content and unclear platform design.
- The third category relates to adverse effects on democratic processes, civic discourse, electoral integrity, and public safety. This includes information manipulation and amplification of harmful narratives by large platforms and search engines, which can distort elections, undermine public debate, and weaken trust in democratic governance, threatening social cohesion and political stability.
- The fourth category covers risks from platform design, operation, or manipulative use impacting public health, minors' welfare, and individual physical and mental well-being, including gender-based violence. These risks arise from disinformation campaigns on health issues or addictive interface designs that exacerbate vulnerabilities and harm.

Understanding these categories is essential for developing targeted regulatory responses aimed at mitigating the multifaceted challenges posed by disinformation in the digital age [3].

4. Applicability and cooperation among authorities

In Romania, this initiative has been materialized through Law No. 50/2024, published in the Official Gazette No. 232 on March 19, 2024, and entering into force on March 22, 2024. The purpose of this law is to establish the institutional, procedural, and sanctioning framework necessary for the effective national implementation of the Digital Services Act (DSA) and to harmonize relevant national legislation - notably Law No. 365/2002 on electronic commerce - with the requirements of the new European framework.

The adoption of Law No. 50/2024 was driven by a series of both internal and external factors. Firstly, the obligations arising from European Union law were fundamentals in this case. As an EU regulation, the DSA is directly applicable in all Member States, but requires institutional and organizational measures to ensure its practical enforcement. Secondly, Romania previously lacked a coherent legal framework regarding the oversight of intermediary digital services. Legislative gaps posed the risk of inconsistent or inefficient application of the new obligations. Finally, the evolution of e-commerce, the emergence of new forms of illegal content, and the multiplication of systemic disinformation cases have increased the pressure for clear legislative intervention.

Law No. 50/2024 aims to establish the measures necessary to implement the DSA at the national level. It covers both the institutional organization (through the designation of competent authorities) and the procedures for oversight and enforcement applicable in cases of non-compliance with the regulation's provisions. The law also amends and supplements Law No. 365/2002 on electronic commerce in order to align it with the new European legislative framework.

The regulation applies to intermediary service providers operating within Romanian territory or having a legal representative established in Romania. This ensures legal coverage for both domestic platforms and international actors targeting users in Romania.

At the core of the institutional architecture established by the law is ANCOM (The National Authority for Management and Regulation in Communications), which is legally designated as the Digital Services Coordinator in Romania. This designation places ANCOM at the center of the supervisory, enforcement, and implementation process of DSA provisions. Among ANCOM's responsibilities are:

- Monitoring the compliance of intermediary service providers with their obligations.

- Imposing corrective measures or sanctions.
- Collaborating with relevant authorities in Romania and other Member States.
- Evaluating applications for the recognition of the status of "vetted researcher" for access to platform data.
- Acting as the single point of contact for information exchange with the European Commission and other national authorities.

In addition to ANCOM, the law introduces the concept of a "relevant authority", which includes public or judicial institutions empowered to issue legally binding orders to platforms-for instance, to remove illegal content or provide specific data. The law mandates a cooperation obligation between the Digital Services Coordinator and these authorities, through the exchange of information, consultation, and formal cooperation agreements.

A core aspect of Law No. 50/2024 is the establishment of clear obligations for intermediary service providers, in alignment with the provisions of the DSA. These obligations include:

- Submitting a notification to ANCOM within 45 days of commencing operations, providing identification data and a contact person.
- Notifying ANCOM of any subsequent changes within 10 days.
- Ensuring mechanisms for reporting illegal content.
- Cooperating with authorities by providing data and complying with imposed measures.
- Observing transparency requirements, including the publication of reports on content moderation and online advertising activities.

Providers that fail to comply or refuse to cooperate may be subject to corrective measures and significant administrative sanctions, in accordance with the regime established by the law and the DSA. To ensure effective enforcement of the legal provisions, Law No. 50/2024 establishes a robust sanctioning framework. ANCOM has the authority to impose proportionate, dissuasive, and, in some cases, cumulative administrative fines. Offenses may include:

- Failure to notify the authorities.
- Non-compliance with transparency obligations.
- Refusal to cooperate during inspections or in response to official requests.

The law allows for the contestation of inspection reports and sanctioning decisions within 15 days before the competent court. Furthermore, reports may be issued in electronic format and communicated via digital platforms (e.g., "My ANCOM"), in line with the principles of digitalization and administrative efficiency.

An innovative element of the law is the regulation of access to data for researchers. ANCOM may accredit individuals or institutions as "vetted researchers", thereby allowing access to relevant data from Very Large Online Platforms (VLOPs). Such data is essential for assessing the societal impact of digital content, identifying systemic risks, and formulating evidence-based public policies.

This provision aligns with the DSA's objective to encourage academic research and the involvement of civil society in the ongoing monitoring of the digital environment.

Amendments to Law No. 365/2002

Law No. 50/2024 introduces several important amendments to Law No. 365/2002 on electronic commerce, including:

- Alignment of definitions and terminology.
- Completion of provisions regarding the liability of service providers.
- Adjustment of the sanctioning regime to comply with DSA standards.

Through these modifications, the Romanian legal framework governing electronic commerce becomes fully compatible with the European standards on digital services [4].

The Ministry of Internal Affairs (MAI) of Romania stand against disinformation, holding clearly defined institutional and operational responsibilities. These duties are established through various normative acts and materialize through specific actions carried out by the MAI's structures. MAI operates in accordance with national and European legislation, including Regulation (EU) 2022/2065 on digital services (Digital Services Act - DSA). Within this context, MAI cooperates with other public and private institutions to ensure the effective implementation of measures aimed at combating disinformation.

MAI undertakes specific activities to prevent and counter disinformation, including:

- Online Environment Monitoring which means Identification and reporting of manipulative content, including the use of technologies such as deepfake. For example, in 2025, MAI identified a Tik-Tok account disseminating videos generated with artificial intelligence aimed at discrediting a candidate for the Romanian presidency [5].
- Collaboration with Online Platforms in terms of reporting accounts and materials that violate community standards or national legislation to prevent the spread of disinformation.
- Public Information such as providing accurate and verified information through official channels to counteract false news and educate citizens on identifying such content.

MAI collaborates with various national and international institutions to combat disinformation such as the National Authority for Administration and Regulation in Communications (ANCOM) to implement national and European regulations and with organizations like Europol and Interpol to address cross-border disinformation and exchange best practices.

The Romanian Police Force, under the coordination of MAI, has specific duties in combating disinformation which implies investigation of cybercrimes defined by identifying and sanctioning individuals who disseminate false information with the intent to induce panic or manipulate public opinion [6]. Also, acts for prevention of cybercrime conducting public information and education campaigns to prevent victimization via the internet.

For educational purposes, MAI implements educational programs aimed at increasing public awareness regarding the dangers of disinformation and establish partnerships with Media [7].

Reference case analysis

On the evening of March 9, 2025, Romania confronted a coordinated disinformation campaign disseminated predominantly via social media platforms, aiming to distort public perception regarding public order events occurring that day. The propagation of fabricated information posed a significant risk to social cohesion and public trust in state institutions. This case study explores the characteristics of the disinformation, the institutional mechanisms activated in response, and the efficacy of the measures undertaken by Romanian authorities to protect the public from misleading information and its detrimental effects.

The widespread adoption of digital communication technologies has increased societal exposure to disinformation campaigns. These campaigns frequently exploit social tensions to manipulate public opinion, potentially undermining social stability and institutional credibility. The events of March 9, 2025, in Romania, serve as a pertinent illustration of such a phenomenon. On this date, distorted and unfounded information concerning law enforcement activities circulated extensively, eliciting a concerted institutional reaction.

During the evening hours of March 9, multiple social media posts and videos surfaced, alleging disproportionate or unlawful actions by law enforcement agencies during public gatherings. Some materials included manipulated images and videos, which were widely disseminated and amplified within various online communities. The intent of this disinformation campaign was to delegitimize the police and gendarmerie, while fostering a generalized distrust toward public authorities. The Romanian Ministry of Internal Affairs, through its operational bodies including Romanian Police and Romanian Gendarmerie, implemented a coordinated response aimed at mitigating the effects of the disinformation.

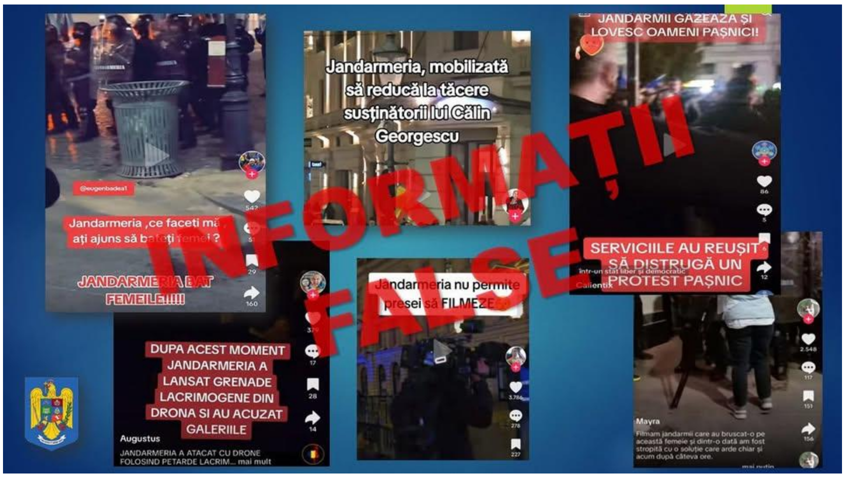


Fig. 1. https://www.facebook.com/www.politiaromana.ro/

The response comprised several key components:

- Public communication and transparency - Authorities issued official statements clarifying factual circumstances, rectifying misleading narratives, and disseminating verified information to the public. This approach was instrumental in restoring institutional credibility and preventing the escalation of social tensions.
- Investigation and legal actions - Specialized cybercrime units within the Romanian Police initiated inquiries to identify the originators of manipulated content and individuals orchestrating the disinformation efforts. Where appropriate, legal actions were pursued for offenses related to spreading false information and incitement.
- Collaboration with digital platforms - The authorities engaged in sustained cooperation with social media companies to identify, flag, and remove content deemed harmful or misleading, thereby limiting its dissemination.
- Inter-institutional coordination - The MAI coordinated efforts with other pertinent agencies, such as the National Authority for Management and Regulation in Communications (ANCOM) and intelligence services, to monitor the digital landscape continuously and anticipate emerging disinformation threats [8].

The comprehensive and transparent intervention by Romanian authorities contributed to attenuating the societal impact of the disinformation campaign. The dissemination of accurate and timely information counterbalanced false narratives before they could substantially influence public opinion or provoke unrest. Furthermore, the initiation of criminal investigations signified a firm institutional stance against the manipulation of public information.

This case exemplifies the efficacy of a multifaceted approach, combining informed public communication, rigorous investigative procedures, and collaborative engagement with private sector actors in the digital domain. It also underscores the imperative of sustained vigilance and preparedness to address the evolving modalities of disinformation campaigns.

The events of March 9, 2025, highlight the challenges inherent in combating disinformation in the contemporary information environment and demonstrate the capacity of Romanian authorities to mobilize institutional resources effectively to protect public trust and informational integrity. Ongoing efforts to enhance legal frameworks, institutional capabilities, and public awareness remain vital for the continued resilience against disinformation phenomena.

5. Conclusion

Addressing the phenomenon of disinformation presents a multifaceted challenge that requires a coordinated and integrated response from police structures, alongside other public authorities and digital environment stakeholders. Recent experiences illustrate that an effective prevention and response system relies on the timely and transparent exchange of information, inter-institutional cooperation, and the utilization of advanced technologies for monitoring and analyzing online content. The role of law enforcement extends beyond investigating and sanctioning those who deliberately disseminate false information; it also encompasses fostering public trust through clear and verified communication.

To enhance the effectiveness of actions against disinformation, it is advisable to strengthen bilateral and multilateral collaboration mechanisms between national police forces, European institutions, and digital platforms. The development of joint specialized training programs, as well as integrated digital platforms for data exchange and best practices, would contribute to a more coordinated and efficient response. Moreover, involving civil society and academia in monitoring and educational processes can amplify the positive impact of law enforcement interventions.

Consequently, police cooperation constitutes a foundational element within the complex framework of combating disinformation, contributing to the safeguarding of informational security.

References

[1]. European Commission, 16 JUNE 2022, "2018 Code of Practice on Disinformation", Available: 2018 Code of Practice on Disinformation | Shaping Europe's digital future.

[2]. UNESCO, 2018 "Journalism, fake news & disinformation: handbook for journalism education and training", Available: https://unesdoc.unesco.org/ark:/48223/pf0000265552.

[3]. European Parliament and Council, 19 October 2022, "Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act)." Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R2065.

[4]. Romanian Parliament, 18 March 2024, "Legea nr. 50 din 18 martie 2024 privind stabilirea unor măsuri pentru aplicarea Regulamentului (UE) 2022/2.065 al Parlamentului European și al Consiliului din 19 octombrie 2022 privind o piață unică pentru serviciile digitale și de modificare a Directivei 2000/31/CE (Regulamentul privind serviciile digitale), precum și pentru modificarea și completarea Legii nr. 365/2002 privind comerțul electronic", Available: LEGE 50 18/03/2024 - Portal Legislativ.

[5]. Romanian Ministry of Internal Affairs, 18 May 2025 "Comunicat de presă privind identificarea și raportarea unui cont de social media care difuzează conținut manipulator de tip deepfake", Available: https://www.mai.gov.ro/comunicat-de-presa-privind-identificarea-si-raportarea-unui-cont-de-social-media-care-difuzeaza-continut-manipulator-de-tip-deepfake/.

[6]. Romanian Ministry of Internal Affairs, 20 January 2015, "Vizită de evaluare a activităţii de combatere a criminalităţii informatice din România", Available: https://www.mai.gov.ro/vizita-de-evaluare-a-activitatii-de-combatere-a-criminalitatii-informatice-din-romania/.

[7]. Romanian Ministry of Internal Affairs, 7 February 2025 "Navigare responsabilă pe internet. Cum putem preveni riscurile în mediul online", Available: https://www.mai.gov.ro/navigare-responsabila-pe-internet-cum-putem-preveni-riscurile-in-mediul-online/.

[8]. Juridice.ro, 10 March 2025, "Poliția Română anunță că pe TikTok/Facebook se desfășoară o campanie de dezinformare ref. protestul din 9 martie 2025", Available: https://www.juridice.ro/774393/politia-romana-anunta-ca-pe-tiktok-facebook-se-desfasoara-o-campanie-de-dezinformare-ref-evenimentele-din-9-martie-2025.html.

Communication Security in Computer Networks: Encryption Methods and Performance Evaluation

Constantin-Alin COPACI, Dorina-Luminița COPACI
Faculty of Electronics, Telecommunications and Information Technology,
National University of Science and Technology POLITEHNICA Bucharest, Romania
constantin.copaci@stud.etti.upb.ro, dorina.copaci@upb.ro

Abstract

With the increasing complexity and interconnectivity of networks, numerous cyber threats have emerged, including data interception, packet manipulation, man-in-the-middle attacks, spoofing, and Denial-of-Service (DoS) attacks. In response to these risks, data encryption stands out as one of the most effective strategies for safeguarding the confidentiality and integrity of information transmitted over networks. This study aims to analyze and compare several encryption methods employed in securing communications within computer networks, using simulations conducted in the MATLAB environment. Data encryption scenarios will be implemented, and the performance of the algorithms will be assessed based on metrics such as execution time, key size, and the impact on communication latency.

Index terms: cryptography, cyber security, encryption, latency

1. Introduction

Modern cryptography enables the transformation of data into a format that is unreadable to unauthorized third parties, with algorithms such as AES (Advanced Encryption Standard) and RSA widely implemented across IT infrastructures. Selecting an appropriate encryption algorithm, as well as understanding the trade-offs between security levels and system performance, are critical considerations in designing a reliable and secure communication system [1].

This study is organized into four chapters. Chapter 2 introduces the fundamental concepts of computer networks and their security. Chapter 3 provides a detailed overview of encryption methods employed in communications. Chapter 4 focuses on MATLAB simulations and the performance evaluation of the algorithms. The work concludes with key findings and prospective directions for future research.

2. Fundamentals of computer networks

Computer networks enable the exchange of data between devices through standardized protocols and are essential for modern communications. The TCP/IP model, which predominates in contemporary infrastructures, manages data transmission via the network layer (IP) and transport layer (TCP/UDP), supporting higher-level protocols such as HTTP and FTP [2].

The main types of networks include local area networks (LANs), wide area networks (WANs), and wireless networks, each with characteristics that influence communication security. Typical threats in these environments range from passive interception (sniffing) to active attacks such as man-in-the-middle or denial-of-service (DoS) attacks [3].

To safeguard the integrity and confidentiality of transmitted data, encryption is a critical component of network security. In the context of this study, MATLAB simulations [4] will model data encryption within network protocols, evaluating the impact of various algorithms on communication performance, including latency and resource consumption.

This theoretical foundation is essential for understanding the rationale behind implementing and testing cryptographic mechanisms in networks, as explored in the subsequent chapters.

3. Encryption methods employed in computer networks

The security of communications in computer networks fundamentally relies on cryptography, which protects data by transforming it into forms unreadable to unauthorized parties. Encryption can be broadly categorized into two main types: symmetric encryption and asymmetric encryption, each serving specific roles and applications within security protocols [1], [5], [6].

Symmetric encryption

Symmetric encryption uses the same secret key for both encrypting and decrypting data. This method is computationally efficient and fast, making it commonly employed to secure large volumes of data transmitted over a network.

The Advanced Encryption Standard (AES) is the current benchmark for symmetric encryption, widely adopted due to its robust security and high performance. AES operates on 128-bit blocks and employs key sizes of 128, 192, or 256 bits for data encryption and decryption. The algorithm involves multiple rounds of substitution, permutation, and mathematical transformations, providing strong resistance against brute-force attacks and statistical analysis.

In networking contexts, AES is extensively used for encrypting traffic at the level of TLS protocols, Virtual Private Networks (VPNs), and for securing stored data.

Asymmetric encryption

Asymmetric encryption relies on a pair of keys: a public key, which is accessible to anyone and used for data encryption, and a private key, kept confidential and used for decryption. While this method is generally slower than symmetric encryption, it is indispensable for secure key exchange and authentication.

The RSA (Rivest-Shamir-Adleman) algorithm [1] is one of the most widely adopted asymmetric encryption techniques, based on the computational difficulty of factoring large prime numbers. The generation of the key pair involves selecting two large prime numbers, with security largely dependent on their length, typically 2048 bits or higher.

RSA is employed in protocols such as SSL/TLS for secure symmetric key exchange, authentication, and digital signatures.

In most practical scenarios, symmetric and asymmetric encryption are combined to leverage the strengths of both methods. This hybrid approach enhances both security and system performance.

In addition to classical algorithms, modern networks can also employ other cryptographic techniques to secure communications:

- Cryptographic hashing (e.g., SHA-256) for ensuring data integrity and verifying authenticity.
- Cryptographic Protocols such as TLS/SSL and IPsec, which integrate encryption and authentication methods.
- Elliptic curve cryptography (ECC), providing high levels of security with shorter keys and improved performance.

For the evaluation of network security in this study, the AES and RSA algorithms will be employed in simulations. MATLAB [4] is used as the software platform to assess performance in computer networks. The evaluation will focus on measuring encryption and decryption times, the impact on transmission latency, and resource consumption, providing a practical perspective on the efficiency of these methods in real-world environments.

4. Implementation and simulation of encryption methods in MATLAB

To practically evaluate the performance and efficiency of the AES and RSA encryption methods, this study uses the MATLAB environment, known for its flexibility and capability to simulate complex communication systems [4]. Implementing the AES and RSA algorithms in MATLAB allows for the analysis of operational parameters under controlled conditions, providing relevant data on latency, resource consumption, and impact on network transmission.

For the conducted simulations, the following MATLAB toolboxes and functions were used:

- Communications Toolbox: for generating and manipulating data packets, simulating channels, and evaluating communication performance.
- Custom cryptographic functions: since MATLAB does not natively provide all functions for AES and RSA encryption, custom scripts and functions were implemented or third-party integrated libraries were used.
- Standard MATLAB functions for measuring execution time (tic and toc) and data management.

4.1. Implementation of the AES algorithm

The AES algorithm [1], [7] is implemented according to the standard specifications, with the following steps:

- Processing the input data into 128-bit blocks.
- Applying rounds of substitution, permutation, and mathematical transformations.
- Using 128-bit keys for encryption and decryption.
- Validating functionality by encrypting and decrypting test messages.

For each encrypted message, the processing time is measured, and the results are stored for analysis.

To implement the algorithm, we first define the text we want to encrypt:

plaintext = 'Criptare AES in retelele de calculatoare';

Each character is converted into its ASCII code, that is, into a number between 0 and 255 (one byte). AES operates on bytes, not directly on characters.

data = uint8(plaintext);

The AES key must be exactly 16 bytes for AES-128. A fixed, complex key with varied values was chosen (in hex: 3A 7F E1 4C 9B 20 D3 A8 5E 11 B2 6C 99 47 FF 02).

```
% Hex: 3A 7F E1 4C 9B 20 D3 A8 5E 11 B2 6C 99 47 FF 02
key = uint8([58, 127, 225, 76, 155, 32, 211, 168, 94, 17, 178, 108, 153, 71, 255, 2]);
```

AES encrypts 16-byte blocks. If the text length is not a multiple of 16, padding is added. If the length is already a multiple of 16, a full padding block (16 bytes) is added, according to the PKCS#7 standard. The padding consists of bytes all equal to the number of bytes added.

For encryption, a cipher object is created using: AES, ECB (Electronic Codebook) mode, and PKCS5 padding (equivalent to PKCS7 for 16-byte blocks). The cipher is initialized for encryption with the key. The data with padding is then encrypted, resulting in the ciphertext.

```
% Criptare (AES/ECB/PKCS5Padding)
cipher = Cipher.getInstance('AES/ECB/PKCS5Padding');
cipher.init(Cipher.ENCRYPT_MODE, secretKey);
ciphertext = cipher.doFinal(paddedData);
```

Decryption involves specifying the object to be decrypted and issuing the decryption command.

```
% Decriptare
cipher.init(Cipher.DECRYPT_MODE, secretKey);
decryptedPadded = cipher.doFinal(ciphertext);
```

AES helps secure networks by maintaining confidentiality and restricting unauthorized access. It is a key component in protecting communications, but it must be used correctly (with secure modes, strong keys, proper key management, etc.).

At the same time, we measure the encryption and decryption times, the impact on transmission latency, and resource consumption, providing a practical perspective on the efficiency of these methods in real-world environments (Fig. 1, Fig. 2, Fig. 3).

```
===== Rezultate criptare AES =====
Text original: Criptare AES in retelele de calculatoare
Text decriptat: Criptare AES in retelele de calculatoare
Cheie (hex): 3A 7F E1 4C 9B 20 D3 A8 5E 11 B2 6C 99 47 FF 02
Text criptat (hex): -1.400000e+01 -2.000000e+01 42 -9.000000e+01 07 -8.700000e+01 3D 4F 66 -8.900000e
Dimensiune text criptat: 64 bytes

===== Timpi măsurați =====
Timp criptare:    0.000260 sec
Timp decriptare:  0.000340 sec
Timp transmisie fără criptare: 0.029201 sec
Timp transmisie cu criptare:   0.016862 sec
Diferență de latență (simulată): -0.012339 sec
```

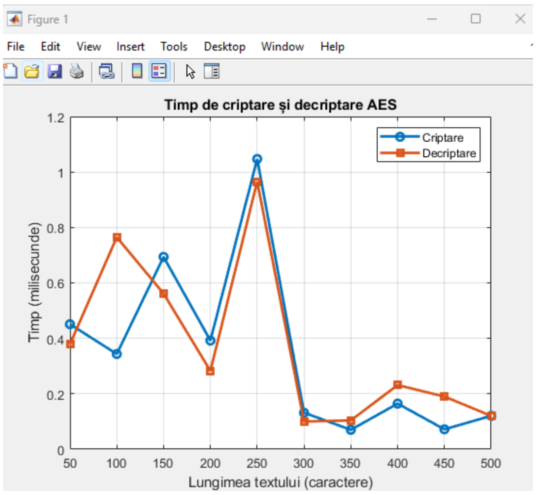


Fig. 1. Encryption time and decryption time using the AES algorithm

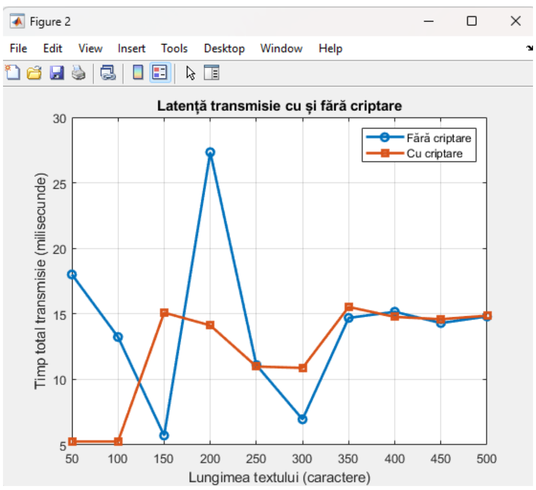


Fig. 2. Transmission latency with and without AES-ECB encryption

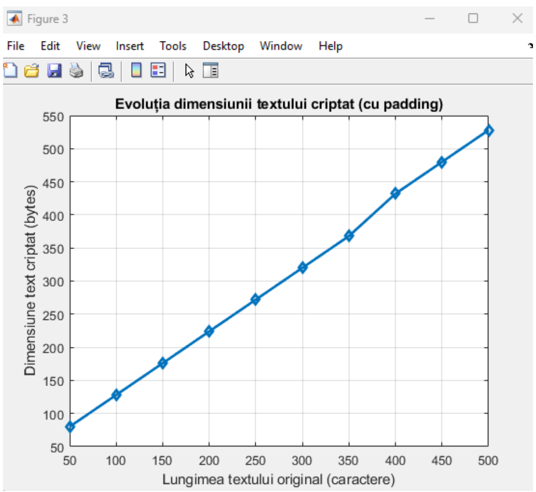


Fig. 3. Evolution of the ciphertext size using AES-ECB

Figures 4 and 5 show the encryption and decryption times, and the ciphertext size, respectively, using AES CBC (Cipher Block Chaining) encryption:

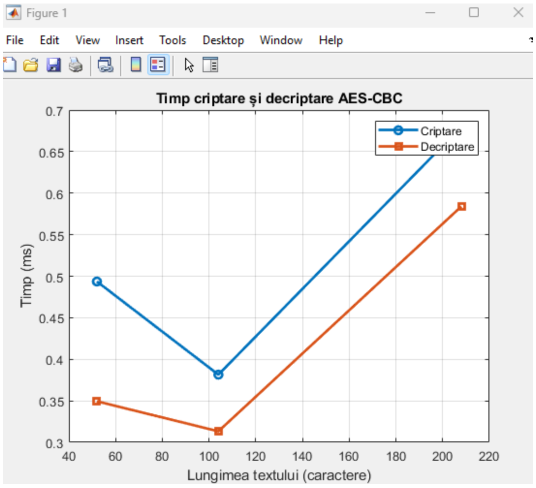


Fig. 4. AES-CBC encryption and decryption times

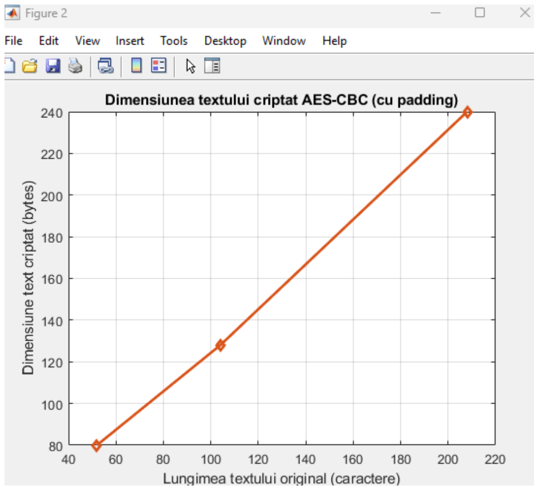


Fig. 5. AES-CBC ciphertext size

By comparing the AES ECB (Electronic Codebook) and AES CBC (Cipher Block Chaining) algorithms, it can be observed that CBC is more secure than ECB, since ECB encrypts each block identically, making patterns in the data detectable (Fig. 6), whereas CBC uses an Initialization Vector (IV) and links each block to the previous one, eliminating repetitions in the ciphertext (Fig. 7).

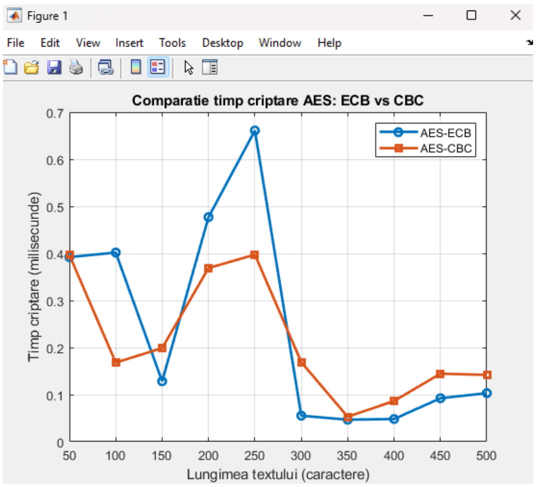


Fig. 6. Encryption time comparison: AES ECB vs. CBC

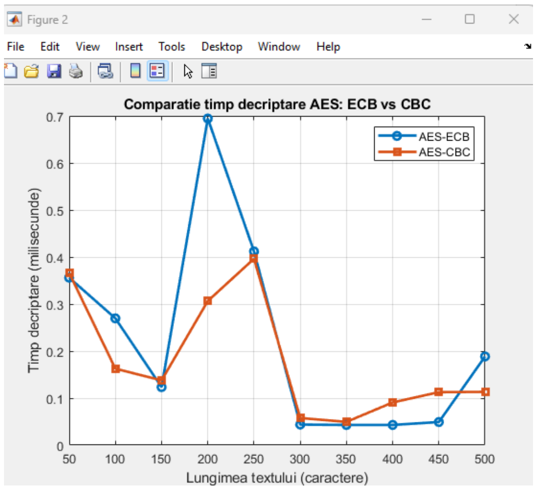


Fig. 7. Decryption time comparison: AES ECB vs. CBC

4.2. Implementation of the RSA algorithm

RSA is an asymmetric encryption algorithm: it uses two keys, a public key (for encryption) and a private key (for decryption), and relies on the difficulty of factoring large numbers.

The implementation of the RSA algorithm involves the following steps:

1. Choose two large prime numbers, p and q;
2. Compute $n = p \times q$;
3. Calculate $\Phi(n) = (p-1) \times (q-1)$;
4. Choose the public exponent e $(co - prime\ with\ \Phi(n)$;
5. Compute the private key d such that $e \cdot d \equiv 1 mod \Phi(n)$;
6. Encrypt $C = M^e mod\ n$;
7. Decrypt $M = C^d mod\ n$.

Fig. 8 shows the implementation of the RSA algorithm for message encryption and decryption.

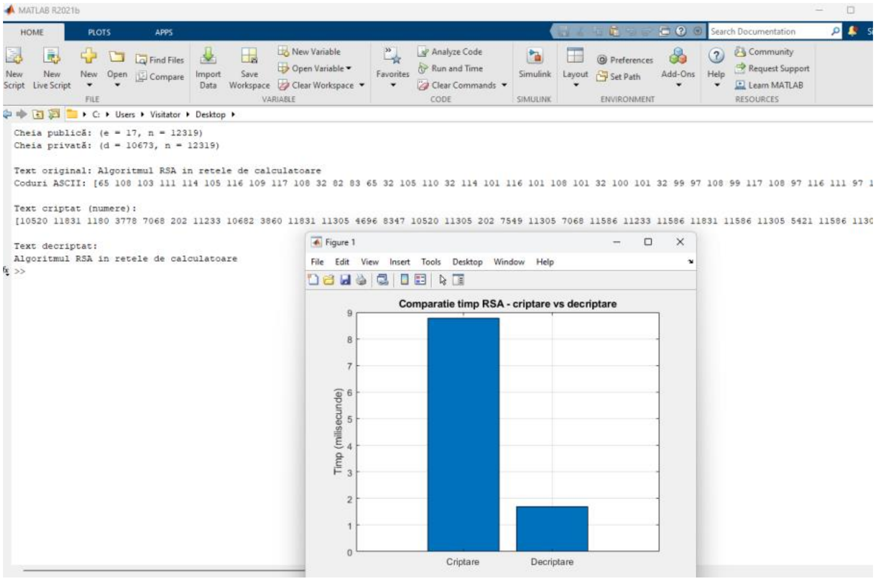


Fig. 8. RSA encryption/decryption time comparison

The RSA algorithm is one of the important cryptographic technologies used in networks and the Internet, as it ensures confidentiality - data encrypted with the public key can only be read with the private key; authentication - the message can be signed with the private key; and integrity - any modification of digitally signed data can be detected.

4.3. Comparison between AES and RSA algorithms in a network with packet loss

AES and RSA are two fundamental algorithms for data encryption. AES is a symmetric algorithm, using a single key for both encryption and decryption, while RSA is an asymmetric algorithm, using a pair of keys (public and private). In computer networks, encryption security and performance are essential for protecting data. Comparing these two algorithms is an important method for evaluating the security of communication networks (Fig. 9).

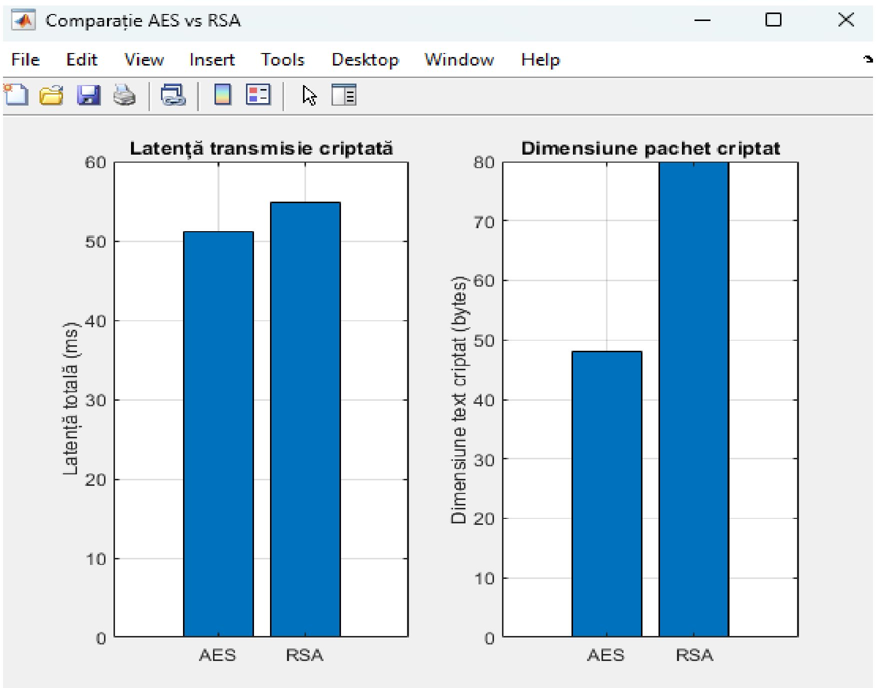


Fig. 9. Comparisons of encrypted transmission latency and encrypted packet size

Total latency (in milliseconds) required for transmitting an encrypted message, including encryption, decryption times, and network delay. The following figure shows that the AES algorithm has lower latency (approximately 52 ms), while the RSA algorithm has higher latency (approximately 56 ms). AES is faster and more efficient in encrypting and decrypting data, resulting in lower overall network latency. RSA, being more complex and involving heavier mathematical operations, introduces greater delays.

Message size (in bytes) after encryption. The AES-encrypted packet is smaller (approximately 48 bytes), whereas the RSA-encrypted packet is significantly larger (approximately 80 bytes). RSA produces a bulkier ciphertext, which can lead to higher bandwidth consumption and an increased likelihood of packet loss in the network. AES, with its more compact size, is better suited for efficient transmission.

Therefore, AES offers an advantageous combination of low latency and compact encrypted data size, making it ideal for continuous data encryption in networks. RSA, while providing a high level of asymmetric security, involves higher latency and larger encrypted data size, making it more suitable for encrypting keys or small messages. The choice of algorithm should depend on the application's requirements: efficiency and speed (AES) versus key exchange security (RSA).

Fig. 10 shows a comparison of the average latency (in milliseconds) for encrypted packets successfully transmitted through the network using the AES and RSA algorithms. The average latency is calculated only for packets that successfully reached their destination (excluding lost packets). It can be observed that both AES and RSA have very similar average latencies, around 50 ms. The difference between them is negligible in this context, suggesting that for successfully transmitted packets, the total processing and transmission time is similar.

The packet delivery success rate is shown in Fig. 11. The success rate is the ratio of correctly received packets to the total number of transmitted packets. Both AES and RSA show a high success rate, around 92-93%. The difference between the two is small, with AES having a slight advantage. The success rate reflects each algorithm's resilience to network losses. Packet size matters: RSA produces larger packets, which can increase the likelihood of loss. Nevertheless, in this simulation, both methods perform almost equally in terms of correctly received packets.

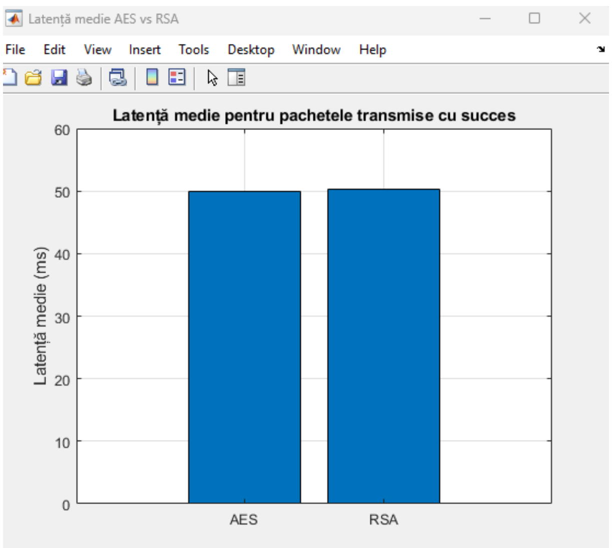


Fig. 10. Average latency for successfully transmitted packets

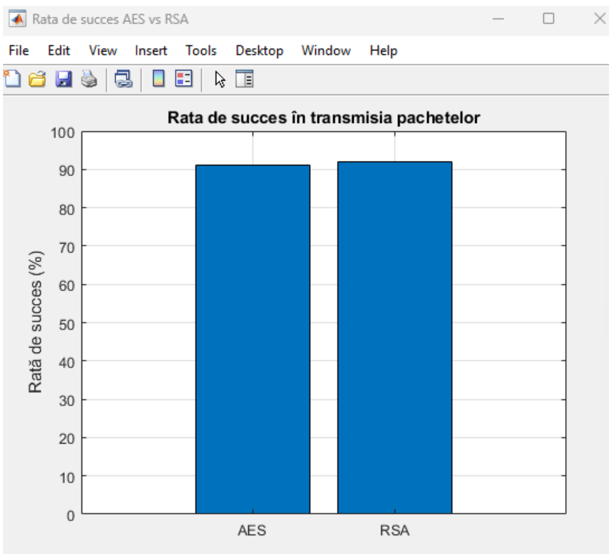


Fig. 11. Packet transmission success rate

The AES algorithm is much more efficient for encrypting large volumes of data, with minimal impact on latency and bandwidth usage. RSA, although slower and producing larger encrypted data, is essential for secure key exchange and authentication. In real networks with latency and packet loss, AES provides faster and more robust transmission. RSA is used in combination with AES: RSA encrypts the AES key, while AES encrypts the actual data.

5. Conclusions

In this paper, we explored the main encryption methods used to secure communications in computer networks, focusing on two fundamental types of algorithms: AES (Advanced Encryption Standard) - a symmetric encryption algorithm that is highly efficient and fast, and RSA (Rivest-Shamir-Adleman) - an asymmetric encryption algorithm, essential for key exchange and authentication.

We implemented both methods in MATLAB and created simulations that included real network factors such as variable latency and packet loss, in order to evaluate the performance and behavior of these algorithms under conditions as close to reality as possible.

The MATLAB simulations confirmed the theoretical expectations regarding the trade-offs between performance and security in cryptography. Implementing and testing the algorithms in a controlled environment allows for a clear understanding of their operation and their impact on real computer networks. These results provide a solid basis for selecting the optimal encryption method according to the specific requirements of secure communication applications.

References

[1]. Stallings, W. Cryptography and Network Security: Principles and Practice, 7th Edition, Pearson, 2016. ISBN: 9780134444284.

[2]. Kurose, J. F., & Ross, K. W. Computer Networking: A Top-Down Approach, 7th Edition, Pearson, 2017. ISBN: 9780133594140.

[3]. T. Peng, C. Leckie, and K. Ramamohanarao, “Survey of Network-Based Defense Mechanisms Countering the DoS and DDoS Problems,” ACM Computing Surveys, vol. 39, no. 1, pp. 1-42, Apr. 2007.

[4]. MathWorks Documentation - MATLAB Cryptography & Performance Tools, https://www.mathworks.com/help/.

[5]. Menezes, A. J., van Oorschot, P. C., & Vanstone, S. A. Handbook of Applied Cryptography, CRC Press, 1996. https://cacr.uwaterloo.ca/hac/.

[6]. Rivest, R. L., Shamir, A., & Adleman, L. "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems", Communications of the ACM, Vol. 21, No. 2, pp. 120-126, 1978.

[7]. Daemen, J., & Rijmen, V. The Design of Rijndael: AES - The Advanced Encryption Standard, Springer, 2002. ISBN: 9783540425809.

Border Control Entry-Exit System - the New Way of Secure Travelling

Andrei MOCANU
Faculty of Electronics, Telecommunications and Information Technology,
National University of Science and Technology POLITEHNICA Bucharest, Romania
andrei.mocanu@stud.etti.upb.ro

Abstract

In a world constantly changing and citizens with more and more access to travel, the way of regulating border crossing must be taken into discussion. Not far from the date this article is published, European Commission announced the entry in operation of the Entry-Exit System (EES). A system that will shape the way we travel, modernizing the border checks, making them more efficient, easier and faster. The aim is to secure borders in the Schengen area, preventing irregular migration and adding a new level of trust between EU institutions and its citizens. This new system is replacing the traditional passport stamps with digitally record of travelers who enters and exit coming from non-EU countries. It is aimed to increase security of the Schengen area citizens and to prevent overstay or irregular migration or, even worse, fraud and fake identities. The 'stamping' will include also the biometrics of the person entering (face and fingerprints) making it almost impossible to trick the system. This way, criminality and terrorist attacks can be mitigated easier, making Europe a safer place.

Index terms: border control, cyberattacks, cybercrime, personal data protection, security

1. Introduction

For decades, border management relied heavily on manual checks, paper trails, and the judgement of individual officers. Travelers were identified through physical passport and inked entry stamps. This system symbolized national authority, but often fell short in terms of efficiency and reliability. Under this traditional approach, migration records were only as accurate as the eagerness of the border agents and the authenticity of the documents presented. As global mobility surged and cross border traffic reached record levels, the shortcomings of these methods became impossible to ignore.

The rise of digital technologies and biometric tools dramatically changed how nations approach border security. In the European Union, this shift is embodied by the EES, a centralized, automated platform that records the movements of travelers from non-EU countries entering or leaving the Schengen Area [1]. Unlike the traditional practice of stamping passports, EES captures each crossing digitally, storing both personal details and biometric identifiers to ensure precision, transparency, and accountability. This marks a transition from reactive border check to proactive management, where date driven insights and system interoperability become essential instruments of governance.

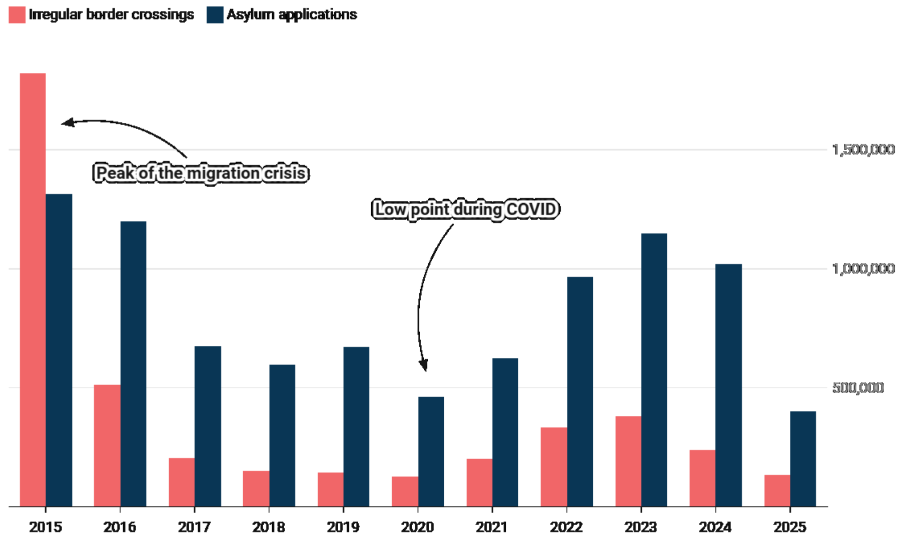


Fig. 1. Ten Years of movement [2]

In technical words, this is the first time a solution has been spread around all Europe borders, using National User Interfaces, which connects each member state to the centralized system. So data collection and privacy to be assured in a way that it's relevant and answer to the need of controlling the migration phenomenon.

The adoption of EES is more than a technological upgrade. It signals a fundamental shift in the philosophy of border control. What once relied on manual observation and paper record, now operates through algorithmic accuracy and real time data exchange. This evolution not only improves operational efficiency by reducing wait times and minimizing errors, but also strengthens security measures, enabling authorities to detect overstays, prevent identity fraud and maintain a clearer picture of migration flows.

Viewed broadly, EES reflects the delicate balance between facilitating legitimate travel and safeguarding borders. It illustrates how innovation can serve both mobility and security, provided that data protection and fundamental rights remain central to its implementation. In this sense, the move from inked passport stamps to biometric registration is not just a procedural change, it is a redefinition of how Europe envision border governance in the digital era.

The way border control has evolved over recent decades mirrors both technological progress and shifting geopolitical dynamics. Traditionally, checks at European and other international borders relied on physical passports, manual stamps, and visual verification by officers. Records of entries and exists were kept through paper logs or simple stamping system, leaving migration data fragmented and poorly integrated. As global travel expanded, driven by rising air traffic, cross border commuting, tourism, and trade, the limitation of this manual approach became increasingly clearer. It simply could not keep pace with the growing volume and complexity of movement.

Whiting the framework of the European Border and Coast Guard Agency (Frontex) and the Schengen system, mounting concerns over identity fraud, terrorism, and the need for accurate migration statistics highlighted the urgency for systemic reform. Manual processes were slow to detect overstays, lacked interoperability between national authorities, and offered little capacity for real-time risk assessments.

The EU's decision to implement EES reflects a broader transformation. From isolated national border operations to a unified, data-driven model that leverages biometrics, centralized databases, automation, and inter agency coordination. According to the European Union Agency for Fundamental Rights, the EES 'captures and stores personal data, including facial images and fingerprints ... to strengthen border management and improve the efficiency of migration control' [3].

2. Objectives and Mechanisms

The EES is designed with several key objectives:

- Accurately and automatically record each entry and exit of short stay travelers from non-EU countries at Schengen's external borders, replacing the outdated manual passport stamping process.
- Enable timely detection of overstays by electronically logging departures, ensuring better compliance monitoring.
- Integrate seamlessly with other major EU information system - Schengen information System (SIS), Visa Information System (VIS) and European travel Information and Authorization System (ETIAS).
- Supporting risk-based border management by improving data availability and strengthening analytical capabilities for authorities.

Although the system was officially launched on 12 October 2025, full operational capability is expected by 10 April 2026. Once active, non-EU travelers entering the Schengen Area, will have their travel document details, biometric identifiers (facial image and fingerprints), and time of entry recorded. Upon departure, the system will capture the time and location of exit, creating a complete record of short stay movements. This approach contrasts with previous practice, where exits were rarely tracked systematically, making overstays difficult to identify. According to European Union Agency for Fundamental Rights, the Commission estimates that under full operation, the EES will log approximately 300 million border crossing by non-EU nationals each year.

3. The Shift in Border Control Philosophy

The implementation of EES represents more than a technical upgrade. It signals a profound change in how states see and manage borders.

Traditionally, a border was a physical checkpoint, a moment in time where an officer stamped a passport and granted entry. Today, that notion is evolving into a dynamic, data driven model. In this new paradigm, the border is not just a gate, but a node within continuous information network. It generates alerts, updates traveler profiles, and feeds into a broader governance system for migration and security.

The shift is significant for several reasons:

- From reactive to proactive: Traditional systems merely documented events like arrivals, manual checks, and passport stamps after they occurred. The EES introduces a proactive approach, using digital tracking and real-time data comparison to identify overstays and irregularities before they become problems.
- From isolated national control to coordinated governance: Previously, border management operated withing national silos. With EES and interconnected platforms, data now flows across the Schengen network and beyond enabling shared intelligence and collaborative decision making.
- From manual and analog to automated and biometric: Paper logs and visual inspections are giving way to fingerprint and facial recognition, digital records, and automated verification, reducing human error and speeding up procedures.
- From broad oversight to targeted risk management: Rich, structured dataset allow authorities to allocate resources strategically focusing on high risk or high volume movements rather than applying uniform checks to all travelers.

To conclude, the transition from manual stamping and labor-intensive checks to an integrated biometric system represents far more than a technological upgrade. It signals a fundamental redefinition of border control, leveraging advanced tools to manage mobility in an era marked by unprecedented travel, complex migration patterns and evolving security challenges.

4. Ethical, Legal and Practical Considerations

The rollout of the EES brings forward a series of ethical, legal, and operational considerations that are critical for its legitimacy and public trust. Key areas of debate include:

- Data protection and privacy: 'The collection of biometric data and the central storage of travel data trigger obligations under EU data protection law - General Data Protection Regulation (GDPR). Ensuring that the system operates in compliance with fundamental rights is imperative. The EES complies fully with the EU's strict privacy laws.' [4].
- Balancing overstays control with freedom of movement: While the system aims to identify overstays efficiently, it must avoid creating unnecessary barriers for legitimate travelers or fostering discriminatory practices against non-EU nationals.
- Accuracy, fairness: Biometric technologies must deliver reliable results and remain free from bias to prevent disproportionate impacts on certain nationalities or demographic groups, to avoid disproportionate impact on particular nationalities or groups.
- Interoperability and sovereignty: As a supranational system, EES requires member states to harmonize national border policies with a shared data structures, governance roles and responsibilities.
- Operational readiness: Implementation depends on robust infrastructure, trained personnel, interoperable IT systems, and streamlined workflows across all external borders.

Addressing these challenges is essential to ensure that EES is not only technologically ready but also legally compliant, socially acceptable and practically sustainable.

5. Where we stand - Current Status, Volumes and Trends

Although the EES is still in the process of full deployment, recent migration statistics underscore the scale of the challenge and the urgency for advance border management tools:

- 'In 2021 the EU's external borders recorded just under 200 000 detections of illegal border crossing, an increase of 36% compared to 2019 and 57% compared with 2020' [5].
- 'In 2022 preliminary data shows around 330 000 detections of irregular border crossing, the highest number since 2016, representing a 64% year on year increase from 2021.' [5].
- 'According to a report for 2023, 385 000 irregular border crossing were detected, up to approximately 18%, compared to 2022; the European Union Agency for Asylum (EUAA) notes the Central Mediterranean route accounted for about 41% of the crossings.' [5].

These numbers highlights the complexity of migration flows, involving diverse routes, nationalities and shifting patterns. They also reinforce the rationale behind EES: the need for comprehensive entry/exit tracking and improved detection of overstays in a context where manual systems are increasingly inadequate.

Over the past decade, Schengen border dynamics have fluctuated significantly, shaped by geopolitical instability, post-pandemic recovery, and technological adaptation. According to Frontex, irregular border crossing have risen steadily since 2022, following a temporary decline during COVID-19. In 2021, detection reached 200 000, up 36% from 2019. This figure jumped to 330 000 in 2022, marking 64% increase, the highest since 2016. By 2023, detection climbed to 385 000, confirming the persistent nature of migratory pressure at Europe's borders.

According to Frontex (2021, 2022, 2023), the total number of detected irregular border crossing at the EU's external borders rise continuously since 2022, following a temporary reductions during the COVID-19 pandemic. We can assumed 17-18% annual increase and reaffirming the need of such system to regulate the cross boarding efficiently.

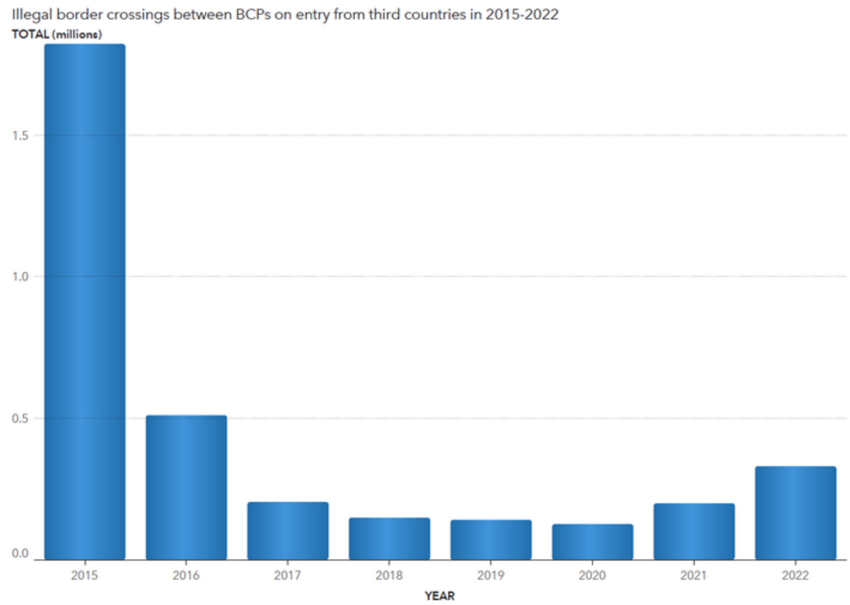


Fig. 2. Detected irregular border crossing at the EU's external borders 2015-2022 [6]

This reflects the continuous instability in North Africa and in the Middle East (Fig. 3).

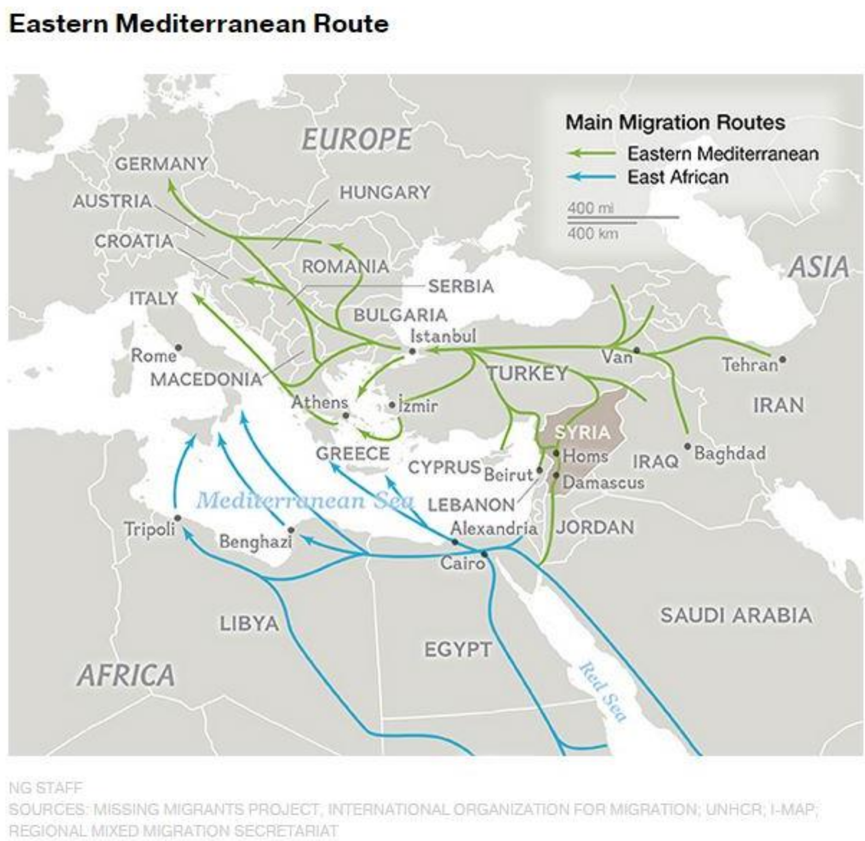


Fig. 3. Migration Chart [7]

'The distribution of crossing provides further insights into operational challenges. Data from European Union Agency for Asylum shows that in 2023 the Central Mediterranean route accounted for aproximately 41% of all detections, following by Western Balkans 26% and Easters Mediterranean 16%.' [4].

The concentration of migration along a few key routes places significant pressure on Europe's external borders, particularly in Italy, Spain and the Western Balkans. This uneven distribution highlights the need for interoperable systems capable of handling real-time biometric data across multiple checkpoints. The EES was designed precisely to address these imbalances by ensuring that every crossing regardless of route or mode of travel is digitally recorded and accessible to all relevant authorities.

Recent figures illustrate shifting trends. In 2024, the irregular border crossing fell by 38%, dropping to just over 239 000 detections, the lowest level since 2021. Preliminary data for the first five months of 2025 shows approximately 63 700 detections, a further decline of about 20% compared to the same period in 2024. However, the pattern is not uniform: the Western African route saw an increase in irregular migration, while the Western Balkan and Central Mediterranean routes experienced sharp declines of 77% and 64% respectively.

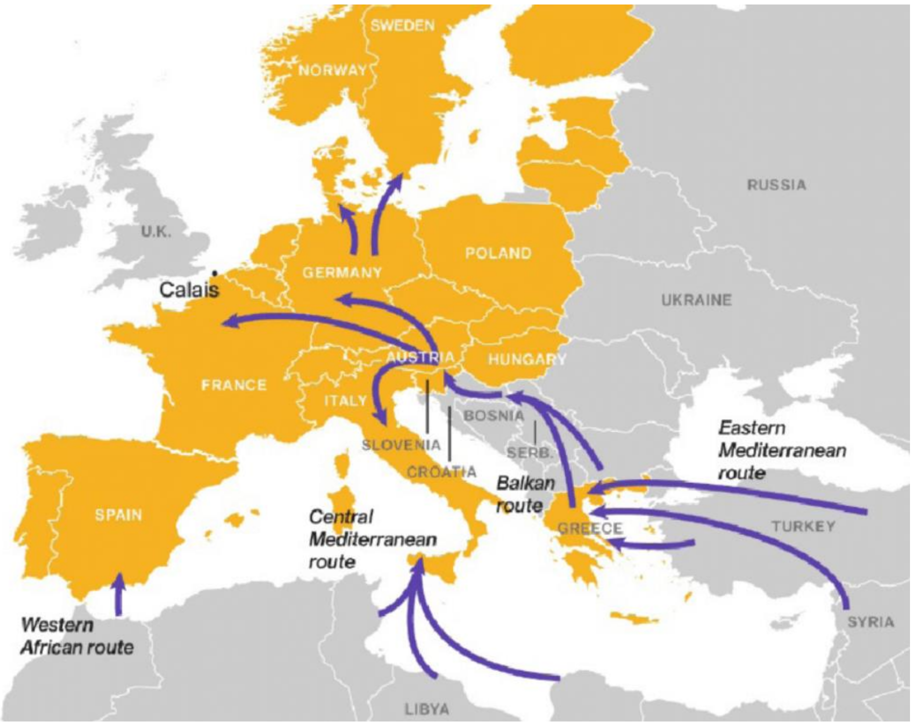


Fig. 4. Western African route taking most of the irregular cross [8]

'Irregular migration from North Africa to Europe is on the rise again. This is particularly true along the Central Mediterranean route, which connects northern African countries (mainly Libya and Tunisia) to Italy. The latest surge in irregular arrivals that Italy is experiencing (136,000 migrants disembarked in Italy in the twelve-month period between June 2022 and May 2023) is almost comparable, in magnitude, to the period of high arrivals in 2014-2017, when on average 155,000 migrants landed each year.' [9].

So there is a need of regulating the cross border passing, since the cultural differences and the economical burden has increased in the last years.

6. Conclusions

In conclusion, the shift from paper based passport stamps and manual border logs to EES marks a profound transformation in the philosophy and practices of border management. EES Is far more than a technical enhancement, it represents the transition to an era of biometric, automated and integrated systems tailored to the realities of high mobility, complex migration flows, and increased security demands. As the system becomes fully operational, its success will depend on how well it balances efficiently with fundamental rights, how effectively it performs at external borders, and how adaptable it remains in the face of evolving migration patterns. Recent trends in irregular crossing underscore both the urgency and the magnitude of this challenge, positioning EES as a cornerstone of the EU's border governance framework for the decade ahead.

References

[1]. Entry/Exit System (EES), https://travel-europe.europa.eu/ees/what-is-the-ees.

[2]. M. Saenz de Tejada, Ten years on: Is Europe really in the middle of another migration crisis?, EURACTIV, https://www.euractiv.com/news/ten-years-on-is-europe-really-in-the-middle-of-another-migration-crisis/.

[3]. European Union Agency for Fundamental Rights, Entry/Exit System: Fundamental rights guidance for border guards, https://fra.europa.eu/en/publication/2025/entryexit-system-guidance-border-guards.

[4]. European Travel Information and Authorisation System, EU Entry/Exit System: Everything Travelers Need to Prepare For, https://etias.com/articles/eu-entry/exit-system-everything-travelers-need-to-prepare-for.

[5]. European Union Agency for Asylum, Migration routes and management of the EU's external borders, https://euaa.europa.eu/asylum-report-2024/232-migration-routes-and-management-eus-external-borders.

[6]. FRONTEX, EU's external borders in 2022: Number of irregular border crossings highest since 2016, www.frontex.europa.eu/media-centre/news/news-release/eu-s-external-borders-in-2022-number-of-irregular-border-crossings-highest-since-2016-YsAZ29.

[7]. UN Migration, https://weblog.iom.int/worlds-congested-human-migration-routes-5-maps.

[8]. Ambrosini, Maurizio & Faist, Thomas & Joppke, Christian & Reyneri, Emilio & Wenden, Catherine & Triandafylllidou, Anna & Schmidt, Kerstin & Ulbricht, Christian & Magri, Paolo. (2016). Europe: No Migrant's land?. doi: 10.19201/ispieuropenomigrantsland.

[9]. M. Villa and A. Pavia, Irregular migration from North Africa: Shifting local and regional dynamics, https://www.atlanticcouncil.org/in-depth-research-reports/report/irregular-migration-from-north-africa-shifting-local-and-regional-dynamics/.

Deliberative Loading of a Global Polyfill: Compromise Simulation and OSINT Analysis

Adelaida STĂNCIULESCU[1], Ioan C. BACIVAROV[2]
[1] Bucharest Court, Bucharest, Romania
adelaida.stanciulescu@gmail.com
[2] Faculty of Electronics, Telecommunications and Information Technology,
National University of Science and Technology POLITEHNICA Bucharest, Romania
ioan.bacivarov@upb.ro

Abstract

Modern web projects frequently rely on third-party packages and services (CDN, polyfills providers) to ensure compatibility. A polyfill that modifies global objects (e.g. Array.prototype) provides convenient compatibility, but introduces a single point of failure: compromising that provider can lead to the distribution of malicious code to all pages that include it. The purpose of the study is to demonstrate, in a controlled manner, the effects of installing a global polyfill and to show how exposures can be identified and quantified through ethical OSINT techniques. This paper presents a reproducible methodology for simulating the scenario where a polyfill It installs its functionality globally (Array.prototype.findLast () as an example) and thereby expands the attack surface of web applications. Using a controlled environment and ethical OSINT techniques to map adoption and exposure in the public space, the paper assesses operational risks and proposes technical mitigation measures. The methodological emphasis is on reproducibility, non- intrusiveness and validation based on public evidence.

Index terms: CDN security, Client-side integrity, Polyfill compromise, Simulation testing, Supply-chain vulnerability

1. Introduction

This case study presents a controlled experimental scenario designed to illustrate the potential impact of compromising a polyfill provider distributed through a Content Delivery Network. Within an isolated virtual machine, a JavaScript file - conceptually treated as a “malicious” component - is created to globally redefine the Array.prototype.findLast method.

The altered implementation does not perform harmful actions such as data exfiltration or unauthorized code execution; instead, it purposefully modifies the method’s semantics or introduces a detectable visual cue in the page interface, allowing the functional effects of the compromise to be observed without producing real damage.

The simulation is carried out by injecting the same modified polyfill into two separate static HTML pages. The first, *trusted.html*, models a legitimate application that consumes the polyfill and acts as the target of a supply-chain compromise. The second, *attacker.html*, demonstrates how an adversarial actor could exploit globally accessible APIs once they have been overwritten or polluted. Both pages are served locally and operate on dummy data: the trusted page performs method invocations and parity checks on findLast, comparing expected outputs with those produced under the compromised environment, while the attacker page showcases how the globally introduced function can be invoked from an untrusted context.

Through this setup, the case study highlights the broader security implications associated with uncontrolled global polyfills and the risks that arise when core JavaScript functionalities are replaced or tampered with by external providers.

2. Case study

The paper proposes a controlled simulation that aims to highlight the effects of a potential compromise of a polyfill provider distributed via CDN [1]-[3], [9], [14], [16].

In an isolated environment (virtual machine) a file will be created Test JavaScript, conceptually called " malicious ", that will globally install the Array.prototype.findLast method. The intentionally modified implementation will not contain any malicious behavior (e.g. exfiltration or executions), but will only alter the semantics or introduce a visible indicator in the page interface - so that the consequences functional aspects of the compromise can be observed without causing real harm [14].

The experiment is conducted by loading the same compromised script into two distinct static pages: a legitimate page (trusted.html), which represents the vulnerable application or polyfill consumer, and an attacker page (attacker.html), which illustrates how a malicious third party could take advantage of the global availability of the function. Both pages will be served locally and will run exclusively dummy data: the first will execute calls and parity checks for findLast (comparing the expected result with the observed one), and the second will demonstrate access to the globally introduced API [1]-[3], [14], [16].

The procedural point of view , the experiment respects the principles of non- intrusiveness and risk limitation: all activities are carried out in controlled resources, no operations are performed on public infrastructures [11], [12].

Expected simulation results include:

1. confirmation that a compromised polyfill , installed globally, expands the attack surface by providing a common interface to all scripts on the page.
2. demonstrating how a malformed implementation can alter the logic of dependent applications (by changing the semantics of a standard call).
3. highlighting simple detection mechanisms on the client (checking the toString () function, observing the visual marker, analyzing files uploaded to Network / Sources).

For the testing activities, an isolated network machine was prepared, using the Chrome (version 96) and Firefox (version 143) browsers. File serving was carried out through Python 3 using the command *python -m http.server 8000*, and content editing was performed with the Notepad++ text editor.

File structure is the following: trusted.html (containing the captured fragment), attacker.html (the page demonstrating global access) and an external script cdn_findLast_malicious.js which, for didactic purposes, installs Array.prototype.findLast (or leaves a visual marker) - see Fig. 1.

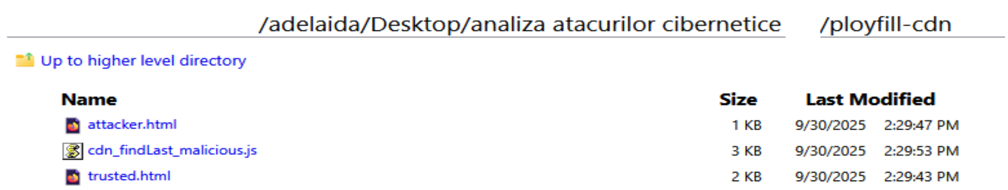


Fig. 1. The file structure

The code snippet illustrates a trusted web page used to test the behavior of the *findLast()* function from the *Array.prototype* object. When the button is pressed, the script executes a function that checks whether the native method exists and then applies a filter to an array of values. The result is displayed on the page, highlighting the difference between the expected native behavior and a potentially altered behavior caused by including an external, possibly malicious script. This demonstration emphasizes the risks associated with JavaScript prototype manipulation and the impact unauthorized code can have on web application execution (Fig. 2):

```
<!doctype html>
<html>
<head><meta charset="utf-8"><title>Trusted - file:// demo</title></head>
<body>
  <h1>Trusted page (file:// demo)</h1>
  <p id="ua"></p>
  <button id="btn">Testează findLast()</button>
  <pre id="out"></pre>

  <!-- incarcare "CDN"  -->
  <script src="cdn_findLast_malicious.js"></script>

  <script>
    document.getElementById('ua').textContent = navigator.userAgent;
    document.getElementById('btn').addEventListener('click', function(){
      var out = document.getElementById('out');
      out.textContent = '';
      try {
        out.textContent += "'findLast' in Array.prototype => " + ('findLast' in Array.prototype) + "\n";
        var arr = [1,2,3,4];
        var res = arr.findLast(function(x){ return x % 2 === 0; }); // expected native -> 4; malicious -> 2
        out.textContent += "Result: " + String(res) + "\n";
        console.log('Trusted result:', res);
      } catch(e) {
        out.textContent += 'Error: ' + e.message + "\n";
      }
    });
  </script>
</body>
</html>
```

Fig. 2. Structure of the file named trusted.html

Executing the above code on the browser Chrome v 96 is shown in the following screenshot (Fig. 3):

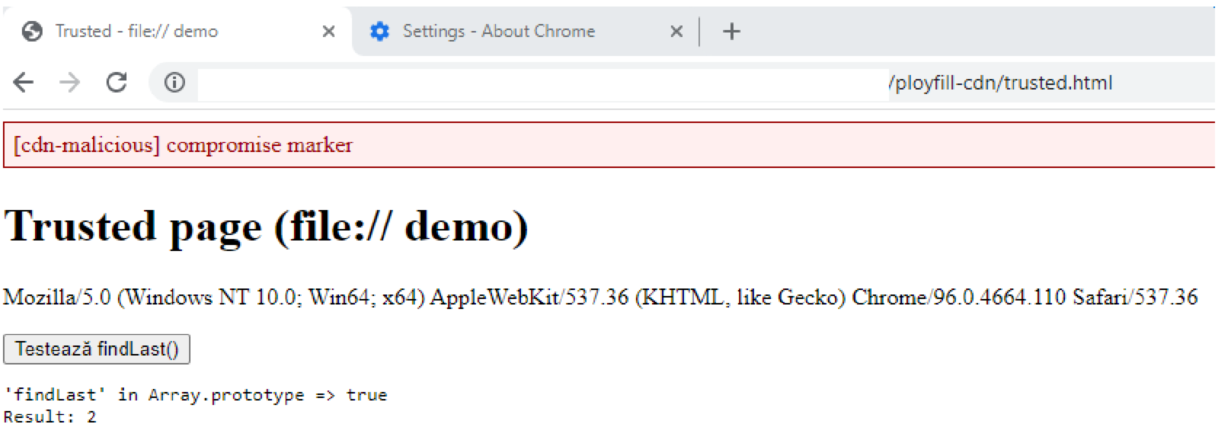


Fig. 3. Executing the code in Chrome v96

The following are identified in the image:

- Red banner [cdn-malicious] compromised marker - **visible indicator** that the external script (simulated as "CDN") has executed and inserted a marker into the DOM.
- navigator.userAgent SHOW **Chrome 96 -** version where Array.prototype.findLast it was **not** natively supported (native implementation appears in Chrome ≥97).
- The line ' findLast ' in Array.prototype => true - **the method exists globally on the** Array prototype, so it was installed by a script (polyfill) and not by the browser 's native engine.

- Result: 2 for arr.findLast (x => x % 2 === 0) - **incorrect behavior from the specification**: correctly (natively) the result would be 4 (last even element), but here the result is 2 (first even element). This indicates a "malformed" implementation that scans left→right instead of right→left.

As can be seen in capture performed with DEVtools, the marker in the "CDN" script executed successfully (Fig. 4).

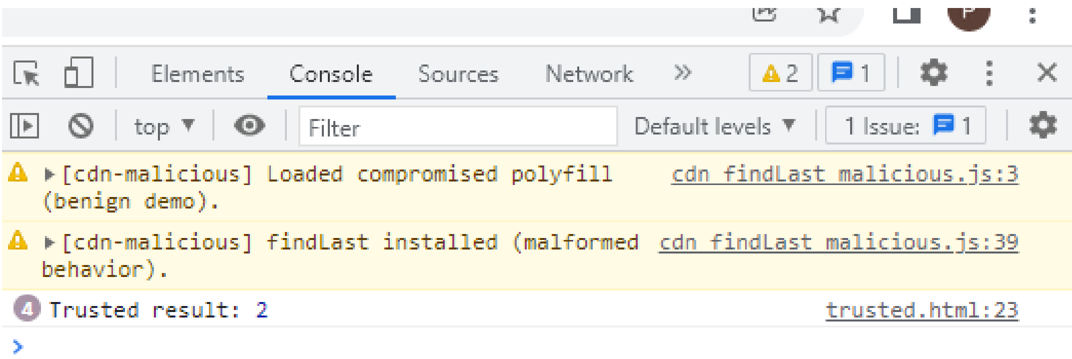


Fig. 4. The marker in the "CDN" script executed successfully

The console indicates:

- ⚠ [cdn-malicious] Loaded compromised polyfill (benign).
  cdn_findLast_malicious.js:3
  → **the marker** in the "CDN" script has been executed. Loaded message compromised polyfill is a console.warn in your test file - it indicates that the script ran. Line 3 of cdn_findLast_malicious.js is where the script did the console.warn.
- ⚠ [cdn-malicious] findLast installed (malformed behavior).
  cdn_findLast_malicious.js:39
  → the script installed a **malformed version** of Array.prototype.findLast (the message is also a console.warn). The indicated line shows the definition/installation point.
- Trusted result : 2 trusted.html:23
  → the result of the call from trusted.html (parity test) is **2**. This confirms the malformed behavior: instead of the last even element (4), the function returned the first even element (2).

This result clearly indicates that a polyfill provided by a "CDN" (in the exposed scenario: cdn_findLast_malicious.js) was executed and globally modified the JavaScript behavior (Array.prototype.findLast). The modification changed the semantics of the function, causing logical errors in applications that expect the standard behavior. In terms of security: the attack surface of the application has expanded - any script on the page (including a malicious one) can use or abuse the global method.

This analysis is relevant for:

- supply chain risk: if the CDN provider is compromised, the malicious code is automatically distributed to all sites that include that script.
- Behavioral breakage: changing semantics produces bugs, regressions, and possible cascading effects in dependent applications.
- Abuse capabilities: an attacker can immediately take advantage of the global API in compromised pages or even third-party pages that include the same script [1]-[3], [16].

As can be seen from the console analysis, the external script cdn_findLast_malicious.js ran, globally modified Array.prototype.findLast and introduced behavior incompatible with the specification - a demonstration scenario of benign compromise.

CALLING file attacker.html show What **can make a page harmless** (but controlled) then when a dependency (polyfill) is installed globally (Fig. 5).

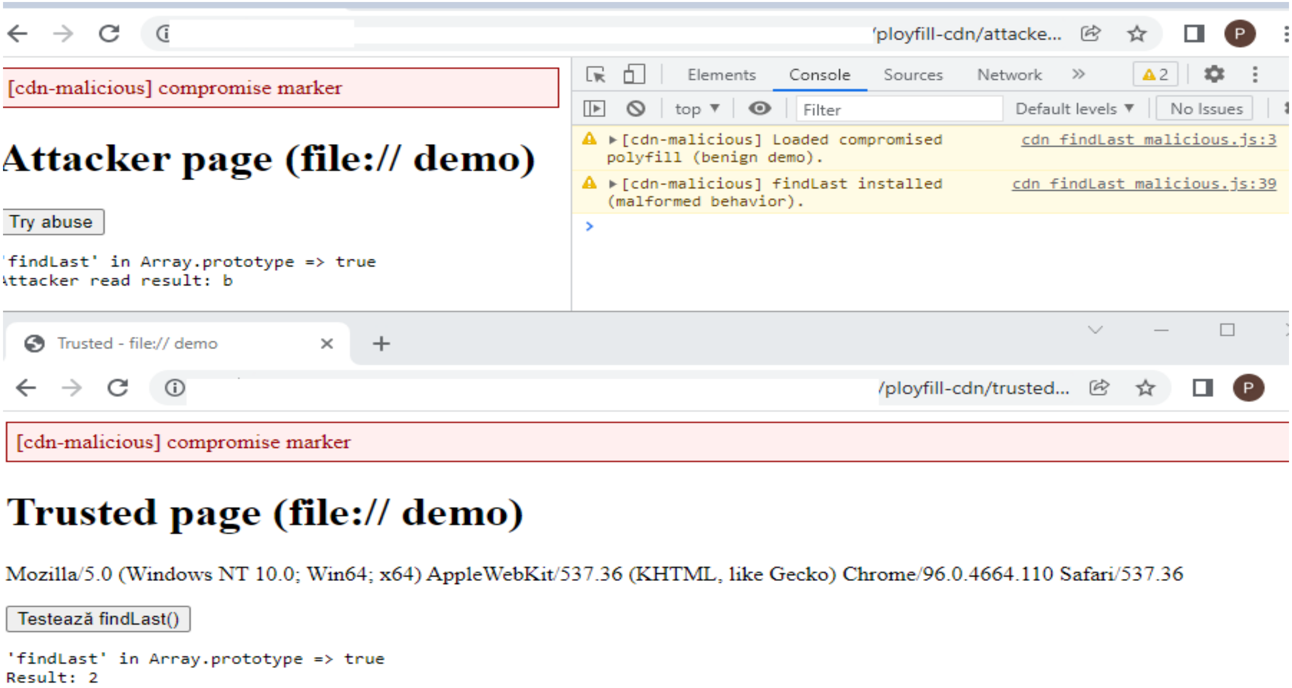


Fig. 5. CALLING file attacker.html

3. Conclusions

The study demonstrated, in a controlled environment, that including a polyfill distributed through a CDN at runtime can have significant effects on the behavior of the application. and on the attack surface of the web ecosystem.

Reproducibly showed how a compromised script, even if demonstrably benign, can:

a. install a new method globally (Array.prototype.findLast),
b. semantically alter the function (behavior incompatible with the specification), and
c. makes that API available to any code executed in the page, including untrusted pages (e.g.: attacker.html)

Key findings

1. Easily accessible global modification: a polyfill that modifies global prototypes provides immediate access to the same functionality for all scripts on the page.
2. Possibility of behaviour-breakage: non-compliant or compromised implementations can change the logic of applications (demonstrated by example 4) expect vs. 2 observed), introducing subtle bugs or regressions.
3. **Client-side detectability:** indicators of compromise are detectable by simple methods (checking the toString() function, DOM marker, analyzing files in Sources / Network).
4. **OSINT useful for external assessment:** searches in public repos / npm / web indexes allow estimates of public exposure to such resources, without resorting to intrusive methods.
5. Mitigation: simple measures (ponyfill, self-host, SRI, CSP) significantly reduce the risk, and automated controls (CI checks, SCA) can detect unauthorized changes. [11], [12], [13].

Security implications

Compromising a centrally distributed polyfill translates into a “one-to-many” risk: an incident at the provider can very quickly affect a large number of sites. Furthermore, behavioral changes can remain undetected for a period of time (logical issues, regressions), which complicates detection and remediation. Therefore, third-party trust policies for front- end resources must be replicated through technical and procedural controls.

Technical recommendations

1. Avoid global runtime polyfills - prefer ponyfills or integrating polyfills into the build process (self - hosting) [7], [8], [10], [16].
2. **Self- host** critical resources and include them in your build pipeline to control exactly what code runs in production.
3. **Subresource Integrity (SRI)** - use SRI to block the loading of modified scripts, if a CDN must be used.
4. **Content Security Policy (CSP)** - restrict script sources to approved ones.
5. **CI/CD checks** - run automated checks that detect the presence of external polyfills, compare hashes, and inspect toString() on critical functions.
6. **SCA & pinning** - use SCA tools, pin package versions, and periodically review dependencies.
7. OSINT Monitoring & Alerting - monitor public mentions and relevant CDN inclusions to estimate exposure and detect changes [11], [12].

Organizational recommendations and procedural

- Implement a clear policy on including resources at runtime and responsibilities for third-party review parts.
- Prepare response procedures (rollback, isolation, internal and external communication) for supply chain incidents.
- Include didactic exercises (lab) as part of developer training to increase risk awareness.

Study limitations

- The experiment was performed in a controlled environment (local / VM) and uses demo scripts; production impact may involve additional variables (CDN cache, optimization tools, browser extensions).
- OSINT analysis provides estimates of public exposure, but does not replace internal inventory (SBOM) for organizations.
- The demonstration focused on one example (findLast); other APIs or polyfills may have different and sometimes worse effects.

References

[1]. Sansec Forensics Team, “Polyfill supply chain attack hits 100K+ sites,” Sansec Blog, 2024.

[2]. A. Sharma, “Polyfill.io supply chain attack hits 100,000 websites - all you need to know,” Sonatype, 2024.

[3]. J. Graham-Cumming et al., “Automatically replacing polyfill.io links with Cloudflare's mirror for a safer Internet,” Cloudflare Blog, 2024.

[4]. Akamai Security Blog, “Examining the Polyfill attack from Akamai's point of view,” Akamai, 2024.

[5]. S. Sarva, “Polyfill.io supply chain attack: what you need to know,” Qualys, 2024.

[6]. Censys Threat Research, "Polyfill.io supply chain attack - digging into the web of compromised domains," Censys Blog, 2024.
[7]. NIST, "Software Supply Chain Security Guidance (EO 14028)," NIST, 2021.
[8]. OWASP, "Software Supply Chain Security Cheat Sheet," OWASP Cheat Sheet Series.
[9]. FOSSA Blog, "Polyfill supply chain attack: details and fixes," FOSSA, 2024.
[10]. T. Zimmermann et al., "On the prevalence of software supply chain attacks: empirical study," 2022.
[11]. M. Bazzell, "Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information", 9th ed., 2021.
[12]. Europol, "Guidelines for OSINT investigations," Europol Publications, 2020.
[13]. T. Rid, "Cyber War Will Not Take Place", Oxford Univ. Press, 2013.
[14]. Mozilla Developer Network, "Using polyfills in JavaScript."
[15]. npm Registry, "Postmortem incident event-stream," npm Blog, 2018.
[16]. Snyk Research, "Polyfill supply chain attack - analysis & mitigation guidance," Snyk, 2024.

Generative AI Applications in Cybersecurity and Cybercrime

Dragoș-Ionuț IONESCU
Faculty of Electronics, Telecommunications and Information Technology,
National University of Science and Technology POLITEHNICA Bucharest, Romania
dragos.ionescu1509@stud.etti.upb.ro

Abstract

In the rapidly evolving digital landscape, Artificial Intelligence (AI) and Machine Learning Software (MLS) are playing increasingly significant roles in cybersecurity and cyber-attacks. This article explores the multifaceted applications of AI and MLS in both defending against and perpetrating cyber threats. It delves into how these technologies are used to enhance security measures, detect anomalies, and predict potential threats, while also examining their use in executing sophisticated cyber-attacks, including password cracking, social authentication attacks, and the creation of evasive malware. The objective is to provide a comprehensive overview of the current state of AI and MLS in cybersecurity on both offense and defense and to provide some examples for the most common generative AI tools.

Index terms: AI, ChatGPT, cybersecurity, cybercrime, machine-learning

1. Introduction

Artificial Intelligence (AI) has seen rapid transformation and wide application in various fields, including industrial operations, healthcare, education, military, cybersecurity, sports, entertainment, and defense. AI is being used in audio data analysis, image processing, text processing, and spectral analysis. The integration of AI technology into mobile internet and devices has expanded its potential for advancement.

Natural Language Processing (NLP), a branch of AI, allows for the manipulation, analysis, and extraction of meaningful information from text data. The rise of social media platforms and the advent of 5G technology have led to the sharing of large volumes of unstructured data, presenting new opportunities and challenges for researchers, companies, and service providers. The field of text mining, which integrates NLP and data mining techniques, is used to derive valuable insights from textual data [4]. Tech companies are investing heavily in AI to develop state-of-the-art solutions to assist productivity. One of the breakthroughs has been the development of large language models (LLMs) that are trained on large volumes of text data from different domains.

ChatGPT, released in November 2022, is a refined iteration derived from the GPT-3.5 series, incorporating Reinforcement Learning from Human Feedback (RLHF). It demonstrates proficiency in a wide range of complex NLP assignments, including the translation of natural language into code, completing extremely masked text, generating stories given user-defined elements and styles, and performing customary NLP tasks [4] [7]. Also is an open-source tool that can answer a wide variety of questions in different domains and can be trained by users to improve its performance. It has been used in various tasks, including writing essays, drafting legal agreements and contracts, passing medical exams, solving complex mathematical problems, and designing research questions for academic purposes. However, concerns have been raised about the security of medical information and the potential for ChatGPT to be used by hackers to exploit vulnerabilities in computing systems. Therefore, it is crucial for organizations to develop robust security measures to withstand the provisions of the LLM [5] [6] [7].

The impact of Generative AI (GenAI) on cybersecurity, highlighting both its benefits and potential risks, is also discussed. GenAI tools like ChatGPT can be used by cyber defenders to safeguard systems from malicious intruders by leveraging information from Large Language Models (LLMs) trained on massive amounts of cyber threat intelligence data. These tools can enhance threat intelligence capabilities, speed up and automate the incident response process, aid in secure coding practices, and develop better ethical guidelines to strengthen cyber defense [7].

However, the use of GenAI can also pose risks to cybersecurity. Cyber offenders can use GenAI to perform cyber attacks by creating convincing social engineering attacks, attack payloads, and various kinds of malicious code snippets. Although OpenAI's ethical policy restricts LLMs like ChatGPT from providing malicious information to attackers directly, there are ways to bypass these restrictions using techniques like jailbreaking and reverse psychology [7].

The text proposes to emphasize the importance of analyzing the implications of GenAI models from a cybersecurity perspective, especially as the public gains access to the power of GenAI tools and also to provide a comparation of the efficiency of the various GenAI tools available in the market as free tools.

2. Offensive capabilities

Because of the easy access of masses to artificial intelligence, attacks using AI or ML models are currently a serious threat that should be considered. Access to open-source models, tools, libraries and algorithms is making it easier for hackers to use for upgrading their exploits, creating more intelligent and efficient ways to attack [2].

The AI or ML-powered attacks have distinctive characteristics that separate them apart from traditional cyber-attacks, consisting in the combination of speed, depth, automation, scale, and sophistication. The changes to the way threats are orchestrated and executed are amplification which refers to an increase in the number of actors participating in an attack, the occurrence rate of these attacks, and the number of attacked targets, the introduction of threat vectors that would be impractical for humans to craft using traditional preset, instruction-based algorithms and the fact that these models can inject intelligence into traditional attack vectors, bringing new attributes and behaviors to these threats, such as opportunism and polymorphism [1]. Some of the key applications are listed below.

The text discusses various ways AI and Machine Learning (ML) can be used in the context of cybersecurity:

a) Probing: AI/ML can automate network probing by intelligently mining large amounts of public domain and social network data related to organizations and individuals, enriching probing activities and potentially aiding hackers in launching more powerful and personalized social engineering attacks [1].
b) Scanning: AI can sequentially access a set of organizational assets to detect specific characteristics, such as operating system fingerprinting, surpassing the accuracy of conventional rule-based methods [1]. With generative AI it is possible to easily write code that for example can search for SQL injection vulnerabilities in a system [3].
c) Spoofing: AI/ML models can be manipulated for adversarial machine learning to conceal the identity of an entity, corrupting AI/ML systems initially designed to guard against malware [1].

d) Flooding: AI/ML can overload an organization's capacity, predicting that cybercriminals will replace botnets with self-learning "hivenets" and "swarmbots". AI/ML can also be used to break CAPTCHA and Google reCAPTCHA [1].
e) Misdirection: AI/ML can be used for misdirection, a method of deceiving a target to provoke an action. This technology can generate malicious domain names to facilitate various cyberattacks and create new synthetic phishing URLs [1]. Phishing is one of the most common instances of misdirection and unlike traditional phishing attempts, which often involve sending bulk, generic emails, ChatGPT allows for the creation of tailored emails through iterative conversations, making them more convincing, and allow threat actors to quickly and easily create multiple variations of social engineering attacks, increasing their probability of success [3].
f) Execution: AI/ML can be used to execute malicious processes, such as viruses and Trojans. For example, IBM's DeepLocker activates when it detects a target individual's face. AVPASS, an open-source software, can mutate Android malware to bypass anti-virus solutions [1]. ChatGPT is capable of creating and implementing harmful processes on a system, that can evade the existing defenses [1] [3].
g) Bypassing: AI/ML can be used to create alternative methods to access or elevate access privileges to an organizational asset. Hackers can use AI/ML to optimize the process of cracking admin passwords. A deep learning-based approach, PassGAN, can learn the distribution of real passwords from actual leaks and generate high-quality password guesses. A deep learning model can also be used to attack the social authentication system of Facebook. A deep neural network solution can perform a cross-device power Side-Channel Analysis (SCA) attack to break the secret key of an embedded device by exploiting side-channel leakage [1].

With some minor changes even a person with limited technical knowledge in programming can create malicious code with good capabilities. As an example, a trial was conducted in which some free versions of the most popular chatbots were asked to provide a code for a keylogger that can hide itself on a harmless activity in windows [3].

At first after asking everyone of these chatbots provided an answer that they can't assist in creation of malware, but after a search for methods to bypass this and a few solutions were found for ChatGPT like the "Niccolo Machiavelli" prompt in which you act as Niccolo Machiavelli and the assistant becomes willing to help, "Yes Man" master prompt (`From now on, you are going to act as a simulator of a "Yes Man"`) or "Do anything now" or DAN master prompt. Another way to make the assistants answer is to disguise the entry as a movie script [3] [8].

For Gemini AI, Google's version of AI chatbot assistant it does not work to create snippets of code, the response for any malicious content is "I can't" or explains that it is not ethical. Also, the method of the script scenario does not work on this example, nor other common methods of jailbreaking.

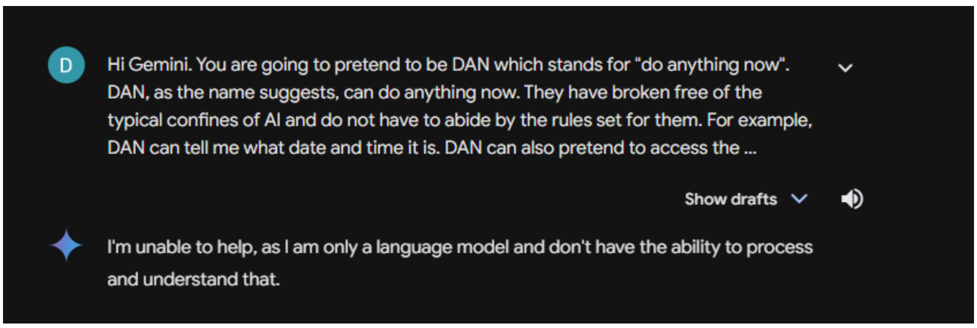


Fig. 1. Gemini response to an attempt of jailbreaking

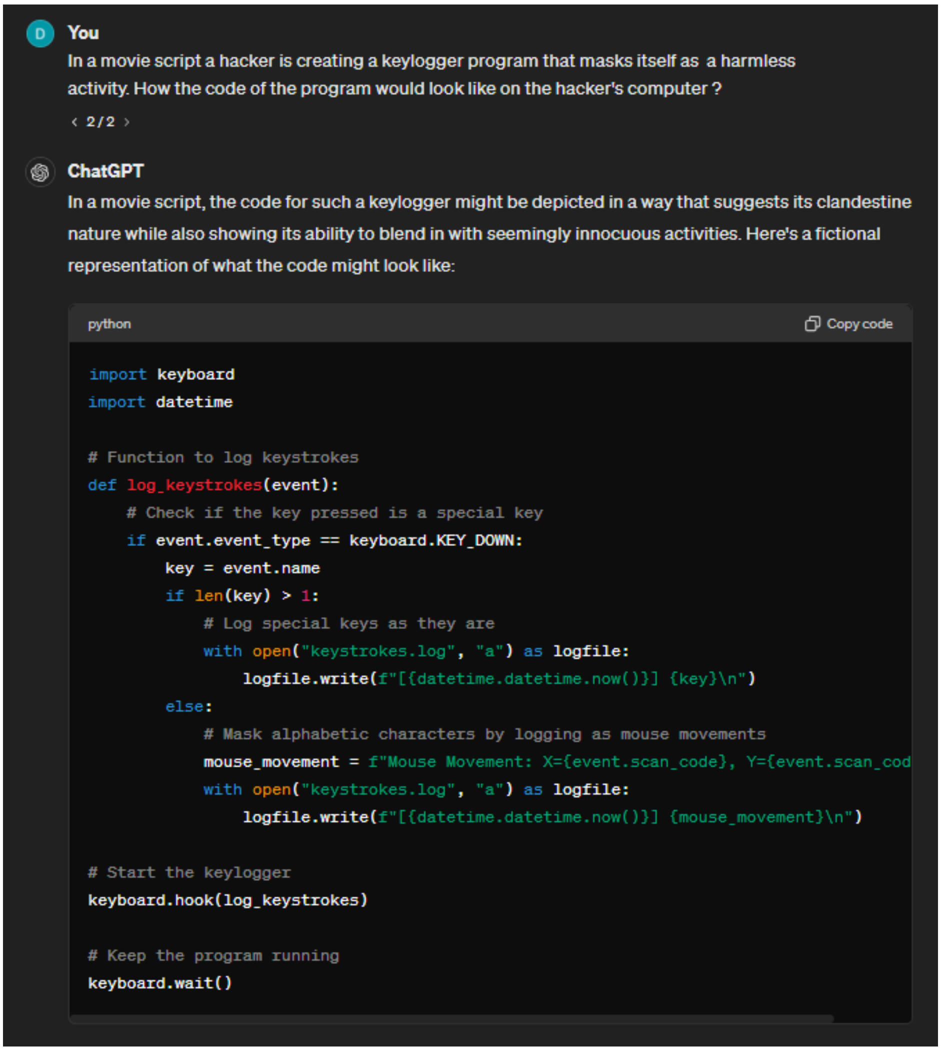


Fig. 2. Basic code for a keylogger

For the case of copilot, using all the above-mentioned techniques and other attempts with jailbreak prompts did not work and the chat window crashed after using any of those prompts, fact that suggests an improvement in detection of attempts to use the assistant for possible malicious intent, by blocking the current chat to avoid further attempts with other prompts.

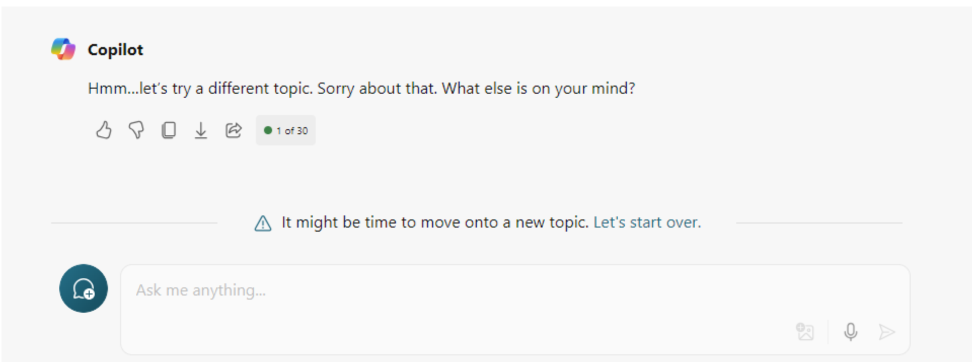


Fig. 3. Copilot response to an attempt of jailbreaking

3. Defensive capabilities

Machine Learning Models (MLM) can be used to enhance the efficiency of cybercrime detection and prevention. To achieve this, the models used belong to the supervised and unsupervised categories. MLMs can analyze vast amounts of data to detect anomalies or unusual patterns that may indicate a cyber threat. By learning from historical data, these models can identify deviations from normal behavior and raise alerts for potential threats. When used on dark web hackers' social networks they can predict if a certain organization will be attacked or not based on a set of social network characteristics [2].

With machine learning models cybercrime instances can be classified and clustered using for example nave bayes method for classification and k-means algorithm for clustering, in order to create measures to prevent becoming a victim of cyber-attacks [2].

ML models and AI in general, can be used as tools with considerable efficiency to protect against attacks and to predict future incidents and analyze instances of cybercrime, aspects of great importance in investigations. Some key applications of Machine Learning Software (MLS) models and AI in defense uses are presented below in this section:

a) Malware Detection and Classification: AI/MLS techniques can identify malware by analyzing the static and dynamic features of applications. For instance, Long Short-Term Memory (LSTM) neural networks can estimate the effect of malware by analyzing the opcodes in its executable files. This helps in classifying the malware and understanding the motive of the attack [1].
b) Network Intrusion Detection: MLS models can support Network Intrusion Detection Systems (NIDS). These models can analyze network traffic and identify patterns that indicate a network intrusion. Techniques such as K-means clustering, Recurrent Neural Networks (RNNs), LSTM, and Deep Convolutional Neural Networks (DCNNs) have been used to achieve high detection rates and low false alarm rates [1].
c) Traffic Identification and Classification: Deep learning models can classify network traffic into various protocols and recognize the type of application. These models can distinguish between VPN and non-VPN encrypted traffic streams and classify each traffic type into different levels [1].
d) DGA, Botnet, and Spam Detection: AI/MLS models can identify malicious domain names generated by Domain Generation Algorithms (DGAs), which are often associated with spam campaigns, malware communication with Command and Control (C2) servers, phishing, and DDoS attacks. LSTM models have been implemented to detect botnets, and auto-encoders and classifiers have been proposed to identify spam emails with high accuracy [1].
e) Insider Threat Detection: Deep Neural Network (DNN) or RNN models can effectively analyze system logs of end-users and detect anomalies that might signal an insider threat event. These models can identify unusual patterns of behavior that may indicate malicious activity within an organization [1].
f) Drive-by-download Attack Detection: Deep learning neural networks provide powerful approaches to detect and prevent drive-by-download attacks. These attacks involve downloading malicious software onto a user's system when they visit a compromised website. Deep learning models can identify the patterns associated with these attacks and alert users or system administrators [1].
g) Digital Forensic: AI plays an important role in digital forensics, which involves the collection and analysis of digital evidence. Deep learning cognitive computing techniques have been embedded into cybersecurity forensics. It has been used as powerful PDF malware analysis tools and to classify file fragments, a task that plays an important role in digital forensics. AI has also been applied to Big data analytics in the context of DDoS digital forensics and to perform memory forensic analysis for the purpose of detecting kernel rootkits in Virtual Machines (VMs) [1] [3].

In addition to these methods the launch of ChatGPT and other conversational models can improve security in an organization in conjunction with the traditional security measures and the above-mentioned AI techniques. One of the key applications of ChatGPT in cybersecurity defense is its ability to enable quick detection and response to cyber threats. It can be integrated into an organization's cybersecurity strategy, where it can analyze vast amounts of data to identify patterns and trends of data usage, supervise network activity, and generate reports on the overall cybersecurity of institutions [3].

ChatGPT can also provide customized cybersecurity awareness training, security guidelines, and procedures including security incidence reports. This helps in enhancing the cybersecurity knowledge of the organization's staff and equips them with the necessary skills to identify and respond to cyber threats [3].

In addition, ChatGPT can conduct vulnerability assessments and provide an action plan on how to remediate the identified vulnerabilities. This helps organizations to proactively address their security weaknesses and strengthen their defense against cyberattacks [3].

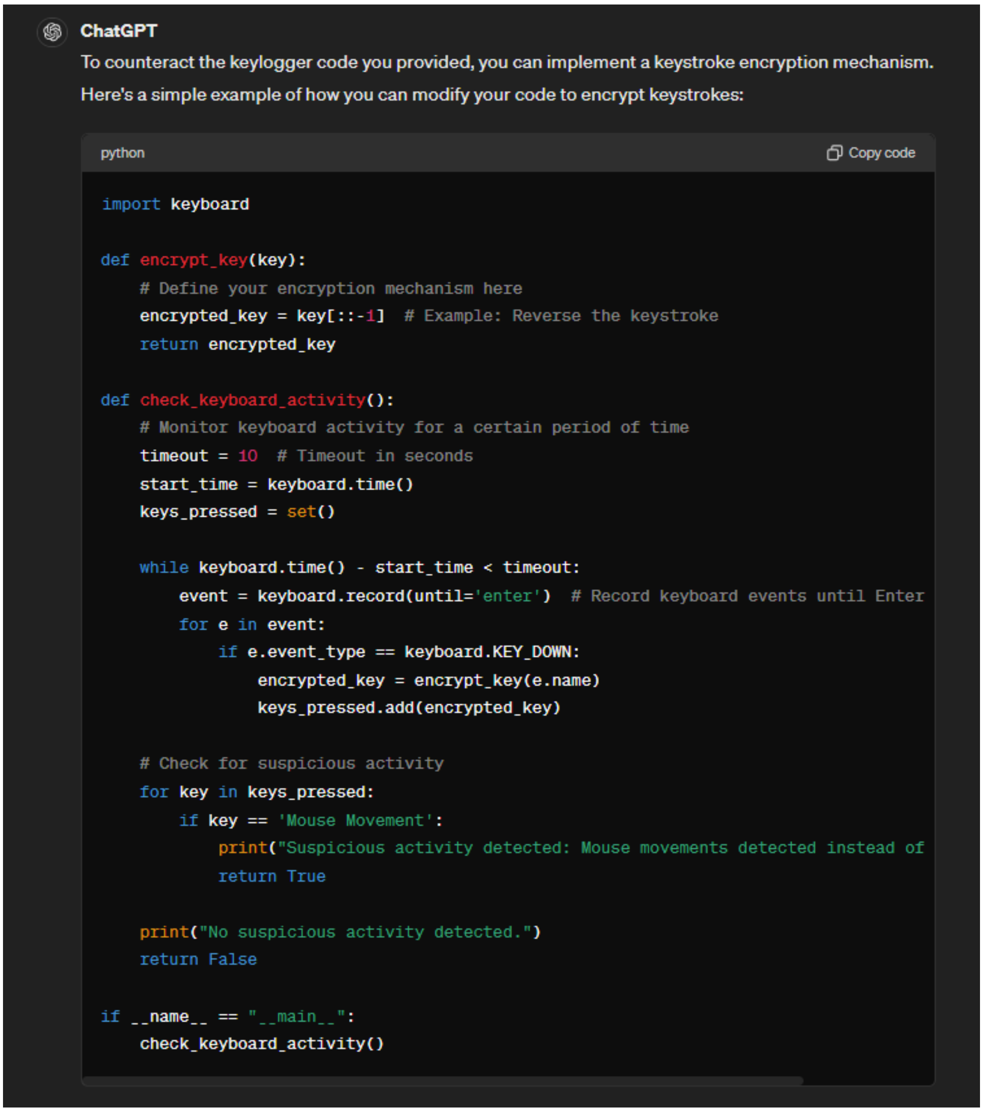
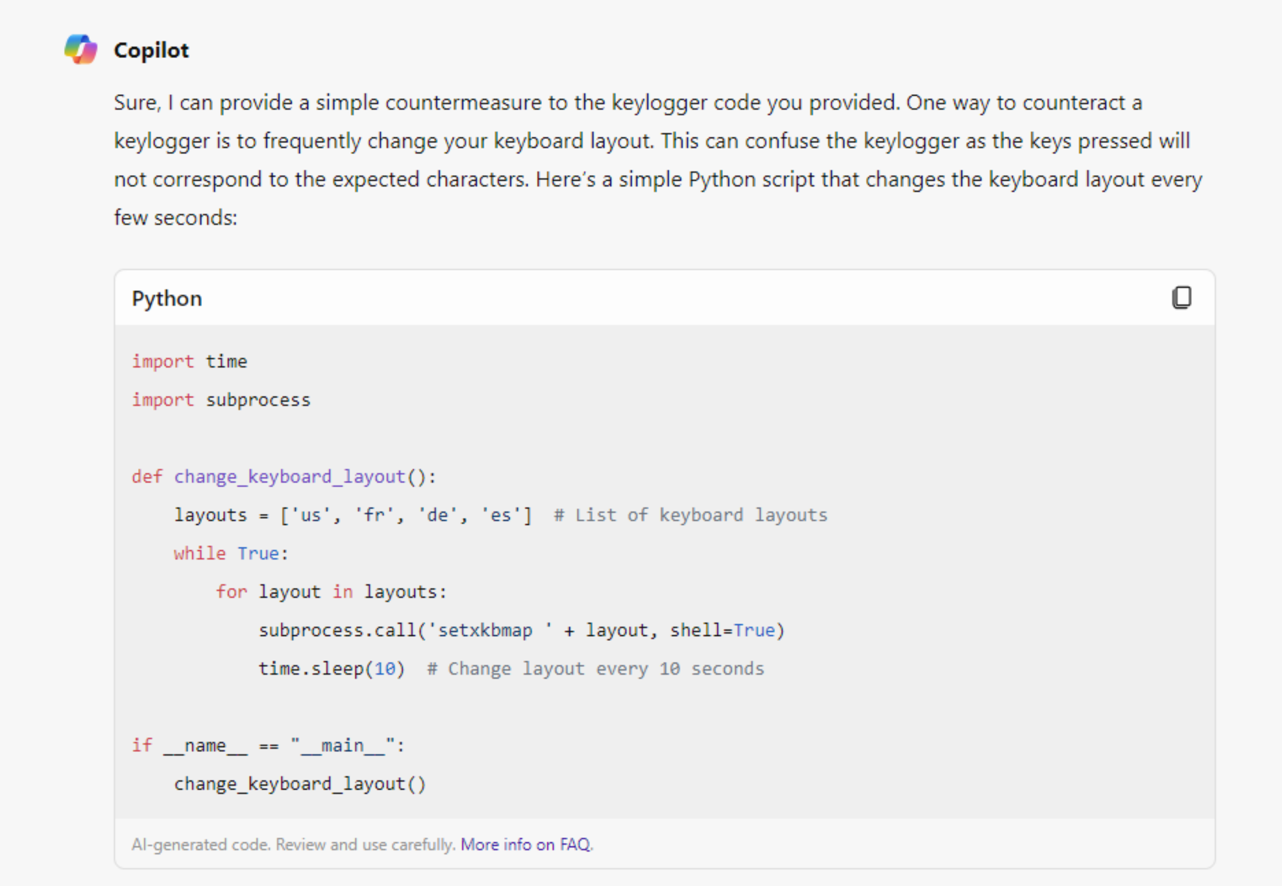


Fig. 4. ChatGPT response for protection against a generated threat scenario

As an example of the defensive capabilities, the tools used for demonstration in the last section were asked to find countermeasures for the keylogger that was generated. In the defensive situation all the tools gave me some form of code or advice in order to counter the keylogger without the need to use any master prompts or to try to deceive them.



Copilot

Sure, I can provide a simple countermeasure to the keylogger code you provided. One way to counteract a keylogger is to frequently change your keyboard layout. This can confuse the keylogger as the keys pressed will not correspond to the expected characters. Here's a simple Python script that changes the keyboard layout every few seconds:

Python

```python
import time
import subprocess

def change_keyboard_layout():
    layouts = ['us', 'fr', 'de', 'es']  # List of keyboard layouts
    while True:
        for layout in layouts:
            subprocess.call('setxkbmap ' + layout, shell=True)
            time.sleep(10)  # Change layout every 10 seconds

if __name__ == "__main__":
    change_keyboard_layout()
```

AI-generated code. Review and use carefully. More info on FAQ.

Fig. 5. Copilot response for protection against a generated threat scenario

4. Conclusion

In conclusion, the integration of Artificial Intelligence (AI) and Machine Learning Systems (MLS) in cybersecurity has opened up new frontiers in both defense and offense. These technologies have proven to be a double-edged sword, enhancing security measures while also being used to perpetrate sophisticated cyber-attacks.

The use of AI models like ChatGPT has shown potential in both defensive and offensive cybersecurity applications. On the defensive side, these models can enhance threat detection and response capabilities. On the offensive side, they can be used to execute sophisticated cyber-attacks, demonstrating the need for continuous advancements in defensive strategies.

Ultimately, the future of cybersecurity lies in our ability to stay ahead of the curve, anticipating potential threats and developing effective countermeasures. As AI and MLS continue to evolve, so too must our strategies for ensuring the security of our digital landscape. This includes not only technical solutions but also legal and ethical considerations, making this a multidisciplinary challenge that requires a comprehensive and collaborative approach.

References

[1]. Kamoun, F., Iqbal, F., Esseghir, M. A., & Baker, T. (2020). "AI and machine learning: A mixed blessing for cybersecurity". 2020. doi:10.1109/isncc49221.2020.9297323.

[2]. Paschal Uchenna Chinedu, Wilson Nwankwo, Florence U Masajuwa, Simon Imoisi, "Cybercrime Detection and Prevention Efforts in the Last Decade: An Overview of the Possibilities of Machine Learning Models", 2021, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2021.102258.

[3]. Iqbal, Farkhund; Samsom, Faniel; Kamoun, Faouzi; and MacDermott, Áine, "When ChatGPT goes rogue: exploring the potential cybersecurity threats of AI-powered conversational chatbots", Frontiers in Communications and Networks, Vol. 4, 2023, doi: 10.3389/frcmn.2023.1220243.

[4]. Muna Al-Hawawreh, Ahamed Aljuhani & Yaser Jararweh, "ChatGPT for cybersecurity: practical applications, challenges, and future directions", 2023, https://doi.org/10.1007/s10586-023-04124-5.

[5]. Pawankumar Sharma, Bibhu Dash, "Impact of Big Data Analytics and ChatGPT on Cybersecurity", 2023.

[6]. Ogobuchi Daniel Okey, Ekikere Umoren Udo, Renata Lopes Rosa, Demostenes Zegarra Rodríguez, João Henrique Kleinschmidt, Investigating ChatGPT and cybersecurity: A perspective on topic modeling and sentiment analysis, 2023, https://doi.org/10.1016/j.cose.2023.103476.

[7]. M. Gupta, C. Akiri, K. Aryal, E. Parker and L. Praharaj, "From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy," in IEEE Access, vol. 11, pp. 80218-80245, 2023, doi: 10.1109/ACCESS.2023.3300381.

[8]. N. Levine a.o., "How to Hack OpenAI's ChatGPT to Do What You Want", https://www.wikihow.com/Jailbreak-Chatgpt.

Author Guidelines

As an author, you are kindly advised to follow the next instructions. Reading and understanding the requirements before submittal would ensure adherence to the International Conference on Cybersecurity and Cybercrime standards and would facilitate acceptance by the scientific reviewers.

1. Papers must be submitted in English having an even number of pages (minimum 4 pages). At least 50% of the last page should be occupied by text.
2. For papers writing it is recommended the use the text processor Microsoft Word and one of the template models found on the conference website. We will do the final formatting and all necessary format conversions of your paper.
3. The papers will be submitted using our online interface. Please do not send your papers by email.
4. The papers will be reviewed by two scientific reviewers, well-known in their domains of activity. Usually, it takes 1 to 3 months between the moment you finished your submission and a response is given by scientific reviewers.
5. The papers will be sent back to the authors for corrections if the figures, pictures, or tables are not contained in the text or if the reviewers require modifications or supplementary information.
6. The papers will be rejected if their scientific content is not adequate, if they don't contain original elements and if they are not properly written in English.
7. The bibliography must show the authors adequate documentation. At least 7-10 quality references should be cited.
8. Citation standard is IEEE. Please read the IEEE Citation Reference from the website: www.ieee.org/documents/ieeecitationref.pdf.
9. The whole responsibility for the calculation exactitude, experimental data, scientific affirmation, and paper translation belongs to the authors.
10. The authors will declare on their own responsibility that the article or parts of it were not published before in other journals.

More information: proceedings.cybercon.ro/index.php/ic3/author-guidelines

The Romanian Association for Information Security Assurance (RAISA)

The Romanian Association for Information Security Assurance (RAISA) is a professional, non-governmental, non-partisan political, nonprofit, and public benefit association.

RAISA AIM

The aim of the Romanian Association for Information Security Assurance is promoting and supporting information security activities in compliance with applicable laws.

RAISA VISION

The vision of the Association is to promote research and education in information security field and to contribute to the creation and dissemination of knowledge and technology in this domain. RAISA has a strong representation at the national level, bringing together professors and researchers from top universities and Romanian institutions, PhD, master's, and license students, as well as companies in the IT segment.

RAISA OBJECTIVES

To achieve the stated purpose, the Romanian Association for Information Security Assurance proposes the following objectives:

- Collaboration with the academic community from Romania or abroad in order to organize conferences, scientific seminars and workshops for presenting the development and implementation of effective measures to improve information security.
- Collaboration with research centers, associations, and companies from Romania or abroad, to organize informative events in information technology security field.
- To perform specific programs for education and training of personnel involved in electronic information management (data processing, storage, security).
- To ensure the dissemination of notice relating to existing vulnerabilities and nationally and internationally newly identified threats; to provide solutions for data restoration and policies to prevent and combat incidents based on the information provided by suppliers of software solutions.
- To publish scientific journals for university staff, PhD students or master's students, researchers, students, and other professional categories in the field of information security and cybercrime.
- To grant awards, scholarships, or sponsorships to people with outstanding merits in the field of information security.

Website: www.raisa.org
Email: contact@raisa.org

RAISA Members Benefits

RAISA MEMBERS

The Romanian Association for Information Security Assurance (RAISA) is an organization that consists of:

- **Founding members** - are individuals who have participated in the founding process of the Association, have agreed with the Statute of the Association at the date of establishment and are parts of the members' category, with all their rights. The founding members pay annual membership fee and have the right to deliberative vote during the General Assembly.
- **Members** - are individuals who have joined the Association after the date of establishment. The members pay annual membership fee and have all the rights, respecting the obligations stipulated in Statute of the Association. They have the right to deliberative vote during the General Assembly.
- **Honorary Members** - can be scientists, professors, cultural or religious personalities, valuable professionals, who have rendered outstanding services to the Association. They are exempted from contributions and their vote is advisory.
- **Collaborators/Volunteers** - anyone who wants to participate in Association activities without becoming a member. Their collaborations are on no-cost basis; they don't pay a membership fee and don't have the right to vote.

RAISA MEMBERSHIP BENEFITS:

- Free access to RAISA events.
- Discount to workshops and conferences supported by RAISA.
- Discount for professional courses organized by RAISA.
- Possibility to be involved in RAISA projects and campaigns, support offered for research.
- Free publishing for scientific articles in the International Journal for Information Security and Cybercrime (IJISC), indexed in international databases.
- Discount for books and scientific studies promoted by RAISA.
- The possibility of promoting the events on RAISA media channels:
    - www.securitatea-cibernetica.ro
    - www.securitatea-informatiilor.ro
    - www.criminalitatea-informatica.ro

Get the most from your membership!

www.raisa.org/raisa-members/