Emerging Cybersecurity Paradigms in Wireless Networks:

Physical Layer Innovation, Trust, and AI-Enhanced Defenses WG1 White Paper (BEiNG-WISE)

Volume editors

Iraklis Symeonidis
Valeria Loscri

This publication is based upon work from COST Action BEiNG-WISE (CA22104), supported by COST (European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career, and innovation.

COST EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

https://cost.eu/

Emerging Cybersecurity Paradigms in Wireless Networks: Physical Layer Innovation, Trust, and AI-Enhanced Defenses WG1 White paper (BEiNG-WISE)

Iraklis Symeonidis - *WG1 Leader*
[1] and Valeria Loscri - *COST Action Chair*[2]

[1]RISE, Research Institutes of Sweden, Stockholm, Sweden
[2]Inria Lille, France

July 23, 2025

Contents

Chapter 1

Executive summary

With the preparation of the 10th Multi-annual Financial Framework (MFF) for the period after 2027, Horizon Europe is developing a renewed work programme whose primary aim is to support the entire innovation cycle-from fundamental research to the commercialization of results. Advancing research from basic science to market deployment demands increased funding, improved research frameworks, stronger collaboration between industry and academia, and effective mechanisms to bridge technological gaps.

The FP10 discussions are unfolding amid growing concerns about Europe's capacity to sustain its global competitiveness. As the EU seeks to boost technological sovereignty and reduce dependence on third countries, FP10 emerges as a critical instrument for strengthening innovation ecosystems and consolidating industrial leadership. In this context, the EU COST Action BEiNG-WISE[1] is investigating both conventional and unconventional wireless cybersecurity solutions that span technical, regulatory, and social dimensions.

As communication technology evolves at an extraordinary pace, its role in relation to end users is also transforming, taking on an increasingly foundational and pervasive character. Next-generation wireless networks are reshaping how people, devices, and critical infrastructure connect, communicate, and operate. The demands for ultra-low latency, massive scalability, and pervasive intelligence are accelerating, while simultaneously introducing new challenges in ensuring trust, security, and resilience.

Emerging technologies – including Artificial Intelligence (AI), novel signal processing methods, quantum computing, and Optical Wireless Communications (OWC) – are converging within the 6G paradigm. This rapidly changing technological reality underscores the necessity of analyzing communication systems from the ground up. A bottom-up perspective enables a detailed understanding of physical architectures and their implications for users from the user and by the user perspective.

In this White Paper, entitled "Emerging Cybersecurity Paradigms in Wireless Networks: Physical Layer Innovation, Trust, and AI-Enhanced Defenses," innovative infrastructures are examined comprehensively, spanning the physical layer to the application layer and the associated services. This first chapter is particularly critical, as the evolution of communication systems has demonstrated that the new gold of the modern era – namely, data – is being generated and exchanged at unprecedented speed and volume. This dynamic environment requires the development of new concepts of trust, a deeper understanding of evolving threats, and the creation of protection mechanisms that prioritize automation through AI systems and the protection of privacy by design.

[1] https://beingwise.eu/publications/deliverables/first-year-deliverable/

1.1 White paper compilation and structure

This White Paper, *Emerging Cybersecurity Paradigms in Wireless Networks: Physical Layer Innovation, Trust, and AI-Enhanced Defenses*, presents 24 state-of-the-art contributions that collectively advance the capabilities of next-generation wireless systems. Reflecting the collaborative efforts of researchers across multiple disciplines, it provides a comprehensive perspective on the opportunities and challenges shaping 5G, 6G, and beyond.

Organized into four thematic chapters, the document explores how physical innovation, trust, cybersecurity, and intelligence converge to create networks that are secure, adaptive, and dependable. By combining technical innovation with policy alignment, regulatory considerations, and adaptive strategies, this White Paper offers an integrated reference for academic researchers and policymakers.

Purpose and Audience This White Paper aims to:

- *Inform* readers about the latest technological developments, emerging threat landscapes, and innovative AI-driven defense strategies.
- *Guide* decision-making in policy, standardization, and implementation by consolidating knowledge on trust frameworks that influence regulatory compliance and support the consideration of risks.
- *Inspire* new research directions at the intersection of infrastructure, security, privacy, and intelligent automation.

Why Read This Document? As wireless ecosystems rapidly evolve, traditional approaches to trust, resilience, and security are no longer sufficient. This White Paper enables the reader to understand:

- How foundational infrastructure advances are expanding both capabilities and attack surfaces.
- Why trust must be dynamic, measurable, and embedded across every layer of the network.
- How AI and Federated Learning (FL) are transforming operational models and threat vectors.
- Which multidisciplinary strategies will be needed to secure hyper-connected environments sustainably.

1.1.1 Chapter Overviews

Infrastructure and Physical Layer Innovation Chapter 2 lays the foundation by examining how emerging technologies – such as terahertz radio interfaces, photonic-based millimeter-wave front-haul, and quantum-compatible satellite links – are converging to create high-performance, ultra-low latency communication infrastructures. It highlights the resulting expansion of the attack surface and introduces energy-efficient security techniques, physical layer protection strategies, and quantum-secure communication methods to reinforce network integrity. Contributions explore topics including:

- Energy-efficient physical layer security for sustainable wireless systems

- Security of optical wireless networks
- Hybrid millimeter-wave photonic links
- Quantum-secure communications

Foundations of Trust and Future Network Integrity Chapter 3 builds upon the technical foundations laid out in Chapter 2, establishing key principles, governance frameworks, and dynamic zero-trust models essential for dependable 6G operations. It examines how trust can be assessed, measured, and operationalized across complex, multi-vendor ecosystems while ensuring privacy and regulatory compliance. The contributions in this chapter address topics including:

- Metrics and methods for trustworthy networks
- Network robustness and resilience measurement
- Trust-aware network management
- The interplay between privacy regulations such as NIS2 and GDPR
- Supply chain cybersecurity

Evolving Threats and Protective Strategies Chapter 4 delves into the dynamic security landscape of next-generation wireless infrastructures. As AI-driven services and pervasive connectivity create new vulnerabilities, the contributions examine both technical and human-centered defenses. It reviews innovations in security protocols, advanced fingerprinting techniques, and vertical-specific challenges in domains such as the Internet of Medical Things and aviation networks. Contributions explore topics including:

- Security protocol innovations and implementation challenges
- Website and RF fingerprinting
- Spoofing attacks in localization
- Human-centric cybersecurity considerations
- IoMT security
- Wireless network security in aviation

Intelligence at the Edge: AI and Federated Learning Focusing on distributed intelligence and privacy-preserving approaches, Chapter 5 investigates how federated learning is redefining scalability and data protection while introducing novel attack surfaces. It discusses the dual role of large language models as both enablers of sophisticated attacks and defensive tools, and highlights the need for energy-efficient, resilient, and deception-aware learning frameworks. Contributions explore topics including:

- Data and model security in communication networks
- Adversarial defense strategies
- Privacy and energy challenges in federated intrusion detection
- Backdoor and poisoning attack mitigation

1.2 Methodology for Collection and Assessment of Contributions

All contributions to this volume were collected through an open call for papers distributed to the participating experts and Working Group 1 members for Cybersecurity in emerging wireless communications of the BEiNG-WISE Cost Action. The evaluation process was designed to be inclusive, constructive, and transparent, focusing on enhancing the quality of each submission rather than exclusion.

1.2.1 Evaluation Instructions and Process

Main Objective The primary goal was to provide constructive feedback that would help the authors improve their contributions.

Submission Type Non-blinded; author names and affiliations were included, and the assigned reviewer was also visible to the authors.

Process: The compilation of this White Paper involved a structured two-round review process to ensure the quality and consistency of all contributions.

Key Aspects Considered

- *Relevance:* Alignment with the Working Group 1 (BEiNG-WISE) objectives outlined in the call for papers.
- *Structure Compliance:* Each contribution was expected to follow the prescribed format: Introduction, State of the Art, Challenges, and Future Work.
- *Content Focus:* Submissions were evaluated based on their ability to present the current state of the art and open questions, rather than solely the authors' own solutions or perspectives.
- *Plagiarism & AI-Generated Content:* Reviewers assessed originality, verified references for credibility, and flagged any significant AI-generated content.

Reviewer-Author Communication Reviewers provided feedback directly to the corresponding authors via email, facilitating iterative refinement throughout both review rounds. All correspondence included the editors in copy for transparency.

1.2.2 Disclaimer

The book editors, chapter editors, and reviewers have undertaken a careful editorial process to ensure that each contribution in this volume reflects the current state of the art in its respective field. Nonetheless, responsibility for the content, accuracy of references, and scientific credibility of each section lies exclusively with its authors. The scientific contributions represent the findings of the respective authors and do not necessarily reflect those of the editors, reviewers, or publisher.

1.3 Acknowledgments

We gratefully acknowledge the valuable contributions of all *Contributors* and *Chapter Editors*. Their insights, dedication, and teamwork were instrumental in shaping this manuscript. This has been a truly collaborative journey, bringing together *24 contributions across four thematic areas.*

We would also like to express our gratitude to the following *Reviewers* for their invaluable efforts in providing constructive academic feedback to the contributors of this White Paper: *An Braeken* (Vrije Universiteit Brussel, Belgium), *Gergely Biczók* (Budapest University of Technology and Economics, Hungary), *Katrien De Moor* (ISCT, NTNU, Norway), *Ana Ferreira* (Faculdade de Medicina da Universidade do Porto, Portugal), *Miranda Harizaj* (Universiteti Politeknik i Tiranës, Albania), *Ali Kara* (Gazi University, Turkey), *Roya Khanzadeh* (Universität Linz, Austria), *Kumar Pardeep* (University of Warwick, UK), *Beatriz Ortega* (Universitat Politècnica de València, Spain), *Mustafa A. Mustafa* (The University of Manchester, UK), *Samaneh Mohammadi* (RISE – Research Institutes of Sweden, Sweden), *Guzin Ulutas* (Karadeniz Technical University, Turkey), *Basak Ozan Ozparlak* (Özyeğin Üniversitesi, Türkiye), *Nicolas Sklavos* (University of Patras, Hellas), and *Stanislav Zvánovec* (Czech Technical University in Prague, Czechia). Their careful reviews and thoughtful insights have significantly strengthened the quality and clarity of this work. We also thank *Sonay Caner-Yıldırım* (Erzincan Binali Yıldırım University, Türkiye) for her valuable contributions to the visual materials and dissemination of this White Paper.

This publication is based upon work from *COST Action BEiNG-WISE (CA22104)*, supported by *COST (European Cooperation in Science and Technology).*

Chapter 2

Infrastructure and Physical Layer Innovation

2.1 Introduction

Chapter Editors: Stanislav Zvánovec [1], Beatriz Ortega [2]

[1] Czech Technical University In Prague, Czech Republic
[2] Universitat Politècnica de València, Spain

The sixth and seventh generation wireless infrastructures aim to integrate new technologies like terahertz radio interfaces, photonic-based millimeter-wave fronthaul lines, and quantum-compatible satellite segments into a single ecosystem. Using such an architecture, a wide range of services requiring sub-millisecond latency can be performed simultaneously, from remote high-precision surgical applications to real-time digital twin scenarios and holographic communications. In the envisaged hyper-connected environment then billions of end nodes will continuously generate data through energy-harvesting sensors. However, the diversification of physical transmission environments and the multi-layering of network topology expand the attack surface, causing new vulnerabilities to emerge at all critical points, from optical components to cloud-based core components. Since classical, layer-limited defense approaches will be insufficient to respond to this speed and complexity, it seems inevitable that security should be considered as an autonomous design dimension that can measure the state of the network in real time, anticipate threats and respond dynamically, and simultaneously consider energy efficiency and quantum-era vulnerabilities.

This section examines breakthroughs in energy-efficient security techniques, advanced signal processing, and quantum-secure communications. Collectively, these innovations reinforce the capacity of future networks to sustain quality, integrity, and scalability in complex environments. Recent studies have shown that the security of the next-generation mobile architecture should be ensured not only by upper-layer protocols but also across the entire physical diversity of the transmission medium. Therefore, the Chapter starts focusing on physical layer security (PLS) strategies for next-generation wireless networks and provides insight on PLS for optical wireless networks. In the following sections, we will delve deeper into the security challenges of millimeter wave-based photonic links and provide an overview of secure quantum communication.

2.2 Energy-Efficient Physical Layer Security Strategies for Sustainable Next-Generation Wireless Systems

Authors: Miranda Harizaj[1], Nazli Tekin[2], Igli Bisha[1]

[1] Polytechnic University of Tirana, Albania
[2] Erciyes University, Türkiye

2.2.1 Introduction

Next-generation wireless systems, encompassing 5G and the emerging 6G technologies, represent a significant leap forward in terms of data rate, latency, and device connectivity. These advancements enable a wide range of applications, from autonomous vehicles and smart cities to industrial IoT and immersive virtual reality. However, the unprecedented scale and complexity of Next-Generation Wireless Systems (NGWS) also introduce new security vulnerabilities, especially at the physical layer. Traditional cryptographic methods, while foundational, may not be sufficient on their own to counter sophisticated eavesdropping and interference threats, particularly those exploiting the broadcast nature of wireless communications [1].

To address these concerns, Physical Layer Security (PLS) has emerged as a promising paradigm, leveraging the characteristics of the wireless channel such as fading, noise, and channel state information to enhance data confidentiality. Techniques like artificial noise injection, beamforming, cooperative jamming, and Intelligent Reflecting Surface (IRS)-assisted communications have shown considerable potential in fortifying wireless transmissions against eavesdroppers [2, 3]. However, these approaches often come at the cost of increased energy consumption, which is especially problematic for battery-powered and energy-harvesting devices that characterize many edge and IoT deployments [4].

Enhancing security while minimizing energy overhead is vital not only for network resilience but also for user experience, particularly in scenarios where battery life is a decisive factor [4]. Recent studies have emphasized the need for adaptive and context-aware PLS strategies that can intelligently manage transmission power, antenna configurations, and security levels based on real-time threat assessment and energy availability [5]. In this context, hybrid solutions that combine lightweight cryptographic primitives with PLS techniques are gaining traction for their ability to balance complexity, energy use, and robustness [6].

Moreover, the integration of energy harvesting technologies such as solar, Radio Frequency (RF), and kinetic energy into secure communication frameworks offers a sustainable pathway for maintaining long-term security operations. These systems can operate continuously with minimal dependence on traditional power sources, making them ideal for remote, unattended, or mission-critical applications [7]. Despite this progress, challenges remain in achieving scalable, real-time, and secure communication in heterogeneous network environments. This paper aims to provide a focused review of the latest advancements in energy-efficient PLS for NGWS, with particular attention to adaptive techniques, IRS-based architectures, and hybrid PLS-cryptographic models. It also highlights the current gaps in practical deployment and proposes directions for future research toward building secure, energy-conscious, and sustainable wireless ecosystems.

2.2.2 State of the Art

The rapid evolution of next-generation wireless systems, including 5G and emerging 6G technologies, has prompted significant research into secure and energy-efficient physical layer solutions. Physical Layer Security (PLS) has emerged as a viable strategy for safeguarding communications against threats such as eavesdropping and jamming by exploiting the randomness of wireless channels. However, the energy demands of conventional PLS techniques such as artificial noise generation, beamforming, and cooperative jamming raise concerns regarding their applicability in energy-constrained environments [4].

To address these concerns, the use of Intelligent Reflecting Surfaces (IRS) has gained prominence. IRSs offer a passive yet programmable means of manipulating wireless propagation environments, thereby enhancing secrecy capacity without the need for active transmission. Studies [8] have illustrated the energy-saving potential of IRS-assisted secure communication systems, making them highly relevant for sustainable Internet of Things (IoT) and low-power networks. Building upon earlier studies, recent work presents a comprehensive Reconfigurable Intelligent Surface (RIS) aided PLS framework specifically designed for 6G-enabled IoT environments, which integrates passive beamforming, artificial noise generation, and cooperative strategies to enhance physical layer security while preserving energy efficiency [9]. This approach further supports the role of RIS as a viable solution for secure, scalable, and low-power IoT deployments within future NGWS. Energy efficiency and security in IoT deployments have also been addressed through cross-layer designs. A recent review [10] highlights multi-layer frameworks that combine lightweight cryptographic measures with adaptive physical layer techniques, enabling robust and sustainable security in ultra-dense device networks.

Additionally, hybrid PLS-cryptographic solutions are being investigated as a means to reduce the computational burden while maintaining high security assurance. These approaches combine lightweight encryption schemes with channel-based security measures to achieve a balance between energy consumption and resilience [11].

The integration of energy harvesting into PLS frameworks is another key trend. Devices powered by ambient sources such as solar, (Radio Frequency) RF, or kinetic energy can maintain continuous security operations without frequent battery replacement. Research has shown that adaptive protocols, which align security levels with real-time energy availability, significantly improve sustainability in low-power deployments [7].

Recent developments also emphasize the role of Fluid Antenna Systems (FAS) in enhancing PLS for Wireless Powered Communication Networks (WPCNs), which offer dynamic adaptation to both external and internal eavesdropping threats while maintaining energy efficiency. It is demonstrated that FAS-equipped users can significantly improve secrecy performance under energy-constrained scenarios [12].

Recent advances also explore the potential of federated learning and Artificial Intelligence (AI)-based optimization to dynamically adjust transmission strategies in response to channel variations and threat conditions. These intelligent, distributed mechanisms support real-time adaptability, a critical requirement for heterogeneous NGWS environments [13].

Table 2.1 summarizes recent contributions in energy-efficient physical layer security for NGWS.

2.2.3 Conclusion – Challenges and Future Work

This White Paper explored current strategies for achieving energy-efficient PLS in NGWS, with a focus on IRS-assisted techniques, energy harvesting, and hybrid PLS-cryptographic approaches. While these solutions offer promising gains in both security and energy efficiency, several challenges remain. These include limited scalability, lack of real-time adaptability, and difficulties in integrating energy-aware PLS strategies into cross-layer architectures. Additionally, many techniques rely on ideal conditions that may not reflect practical deployment environments.

Table 2.1: Recent Contributions to Energy-Efficient Physical Layer Security in NGWS.

Paper	Focus Area	Key Technique	Energy Efficient	Application Context
Wei et al. (2020) [4]	Conventional PLS	Artificial Noise, Beamforming, Jamming	✗	General Wireless Networks
Ihsan et al. (2022) [8]	IRS-assisted PLS	Passive Beamforming via IRS	✓	Sustainable IoT
Xing et al. (2024) [9]	RIS-enhanced PLS	Passive Beamforming, Cooperative Jamming, Artificial Noise	✓	6G-enabled IoT
Mustafa et al. (2024) [10]	Cross-Layer Security	Lightweight Crypto + Adaptive PLS	✓	Dense IoT Networks
Popoola et al. (2024) [11]	Hybrid Security	Lightweight Encryption + Channel-Based PLS	✓	Smart Home Healthcare
Pan et al. (2021) [7]	Energy Harvesting + PLS	Differential Privacy + IRS-Aided Harvesting	✓	6G IoT with Ambient Energy
Ghadi et al. (2025) [12]	FAS-aided PLS	Beamforming with Fluid Antennas	✓	Wireless Powered Networks
Hu et al. (2025) [13]	AI-Optimized PLS	Federated Learning + Model Segmentation	✓	Heterogeneous Edge IoT

Future efforts should prioritize the development of standardized, scalable frameworks for secure NGWS that can operate under resource-constrained conditions. AI-driven adaptive control, cross-layer integration, and testbed validation are key to ensuring that energy-efficient PLS solutions can be translated from theory to real world impact. Addressing these gaps will be crucial to building resilient, secure, and sustainable 6G and IoT infrastructures.

2.2.4 References

[1] Naeem, F., Ali, M., Kaddoum, G., Huang, C., and Yuen, C., "Security and privacy for reconfigurable intelligent surface in 6G: A review of prospective applications and challenges," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1196–1217, 2023. Available: https://ieeexplore.ieee.org/document/10121733. doi:10.1109/OJCOMS.2023.3273507.

[2] Wang, H.M., Bai, J., and Dong, L., "Intelligent Reflecting Surfaces Assisted Secure Transmission Without Eavesdropper's CSI," *IEEE Signal Processing Letters*, vol. 27, pp. 1300–1304, 2020. Available: https://ieeexplore.ieee.org/document/9146177. doi:10.1109/LSP.2020.3010170.

[3] Hong, S., Pan, C., Ren, H., Wang, K., and Nallanathan, A., "Artificial-Noise-Aided Secure MIMO Wireless Communications via Intelligent Reflecting Surface," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7851–7866, 2020. Available: https://ieeexplore.ieee.org/document/9201173. doi:10.1109/TCOMM.2020.3024621.

[4] Wei, Z., Masouros, C., Liu, F., Chatzinotas, S., and Ottersten, B., "Energy- and cost-efficient physical layer security in the era of IoT: The role of interference," *IEEE Communications Magazine*, vol. 58, no. 4, pp. 81–87, 2020. Available: https://ieeexplore.ieee.org/document/9071996. doi:10.1109/MCOM.001.1900716.

[5] Duan, Z., Chang, Z., Xie, N., Sun, W., and Niyato, D., "Adaptive Strategies in Enhancing Physical Layer Security: A Comprehensive Survey," *ACM Computing Surveys*, vol. 57, no. 7, pp. 1–36, 2025. Available: https://dl.acm.org/doi/10.1145/3715319. doi:10.1145/3715319.

[6] Garg, D., Rani, S., Herencsar, N., Verma, S., Woźniak, M., and Ijaz, M.F., "Hybrid Technique for Cyber-Physical Security in Cloud-Based Smart Industries," *Sensors*, vol. 22, no. 12, p. 4630, 2022. Available: https://www.mdpi.com/1424-8220/22/12/4630. doi:10.3390/s22124630.

[7] Pan, Q., Wu, J., Zheng, X., Yang, W., and Li, J., "Differential Privacy and IRS Empowered Intelligent Energy Harvesting for 6G Internet of Things," *IEEE Internet of Things Journal*, vol. 9, no. 22, pp. 22109–22122, 2021. Available: https://ieeexplore.ieee.org/document/9514552. doi:10.1109/JIOT.2021.3104833.

[8] Ihsan, A., Chen, W., Asif, M., Khan, W.U., Wu, Q., and Li, J., "Energy-Efficient IRS-Aided NOMA Beamforming for 6G Wireless Communications," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 4, pp. 1945–1956, 2022. Available: https://ieeexplore.ieee.org/document/9903905. doi:10.1109/TGCN.2022.3209617.

[9] Xing, Z., Wang, R., and Yuan, X., "Reconfigurable Intelligent Surface Aided Physical-Layer Security Enhancement in Integrated Sensing and Communication Systems," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 4, pp. 5179–5196, 2024. Available: https://ieeexplore.ieee.org/document/10306322. doi:10.1109/TVT.2023.3329992.

[10] Mustafa, R., Sarkar, N.I., Mohaghegh, M., and Pervez, S., "A Cross-Layer Secure and Energy-Efficient Framework for the Internet of Things: A Comprehensive Survey," *Sensors*, vol. 24, no. 22, 2024. Available: https://www.mdpi.com/1424-8220/24/22/7209. doi:10.3390/s24227209.

[11] Popoola, O., Rodrigues, M.A., Marchang, J., Shenfield, A., Ikpehai, A., and Popoola, J., "An optimized hybrid encryption framework for smart home healthcare: Ensuring data confidentiality and security," *Internet of Things*, vol. 27, p. 101314, 2024. Available: https://doi.org/10.1016/j.iot.2024.101314. doi:10.1016/j.iot.2024.101314.

[12] Ghadi, M., Kaveh, K., Wong, K., Martin, D., Jantti, R., and Yan, Z., "Physical Layer Security in FAS-Aided Wireless Powered NOMA Systems," arXiv preprint arXiv:2501.09106, Jan. 2025. Available: https://arxiv.org/abs/2501.09106.

[13] Hu, M., Zhang, J., Wang, X., Liu, S., and Lin, Z., "Accelerating Federated Learning With Model Segmentation for Edge Networks," *IEEE Transactions on Green Communications and Networking*, vol. 9, pp. 242–254, 2025. Available: https://doi.org/10.1109/TGCN.2024.3424552. doi:10.1109/TGCN.2024.3424552.

2.3 Physical Layer Security in Optical Wireless Communications

Authors: S. Zvanovec[1], S. R. Teli[1], C. Guerra Yanez[1], M. Petkovic[2], N. Stevens[3], L.N. Alves[4], M. Biagi[5], Z. Ghassemlooy[6], I. Enesi[7], M. Harizaj[7], R. Qafa[8], and V. Loscri[9]

[1]Czech Technical University in Prague, Czech Republic
[2]University of Novi Sad, Faculty of Technical Sciences, Serbia
[3]KU Leuven, Belgium
[4]Instituto de Telecomunicações Aveiro, Aveiro, Portugal
[5]Sapienza University of Rome, Rome, Italy
[6]Northumbria University, Newcastle Upon Tyne, United Kingdom
[7]Polytechnic University of Tirana, Albania
[8]Innovation Center for Security and Defense, Albania
[9]Inria Lille, France

2.3.1 Introduction

Wireless-based Internet of Things (IoT) assists in developing and implementing smart indoor environments (i.e., homes, offices, hospitals, and industries) by interconnecting smart devices and humans, as part of the 5th generation (5G) and beyond wireless networks. Optical wireless communication (OWC) has emerged as a complementary technology or an alternative to existing radio frequency (RF)-based wireless networks to meet the growing traffic demands at many levels from indoor to outdoor applications. In indoor environments, the OWC technology can use both the visible and infrared light spectrums, offering a very broad bandwidth for a range of applications. Light emitting diode (LED)-based lighting fixtures together with photodetectors and image sensors (i.e., cameras), commonly referred to as visible light communications (VLC) and optical camera communications (OCC), can be used to establish wireless links in IoT devices and applications. It is possible to achieve a secured network by ensuring that the availability, integrity, and confidentiality (AIC) triad is satisfied, however, networks that contain IoT nodes, can frequently be subject to breaches of security. Though OWC offers inherent security, there are still several vulnerability issues. In general, the security requirements for OWC are the same as those for other wireless networks, that is to protect data being transmitted from attacks such as eavesdropping, denial of service or jamming, node compromise attacks, etc. Indoor VLC broadcasts data within limited space in rooms however wide beams and reflected and scattered light can be received by eavesdroppers (Eve). Outdoor OWC can be detected by Eve being close to the receiver or in case of scattered light even off the beam propagation, systems can be jammed and interfered with.

To ensure secure transmission of data over wireless channels, classical methods employ key-based cryptography for end-to-end encryption and decryption while the physical layer (PHY) is responsible for ensuring the reliability of the communication links between entities in a shared collision domain, e.g., two terminals in point-to-point links or terminals within wireless local area networks. By utilizing the physical channel, PLS limits the amount of information that can be extracted at the bit level by the eavesdropper. Typically, the PHY ensures reliable communication between legitimate users, while the upper layers, i.e., the network layer, protect and secure the data. However, with the expansion of broadcast networks, new techniques leveraging the PHY have emerged to enhance secure communication. PLS has emerged as a promising approach to limit the information accessible to eavesdroppers by exploiting the randomness of noise, channel state information (CSI), and various resources,

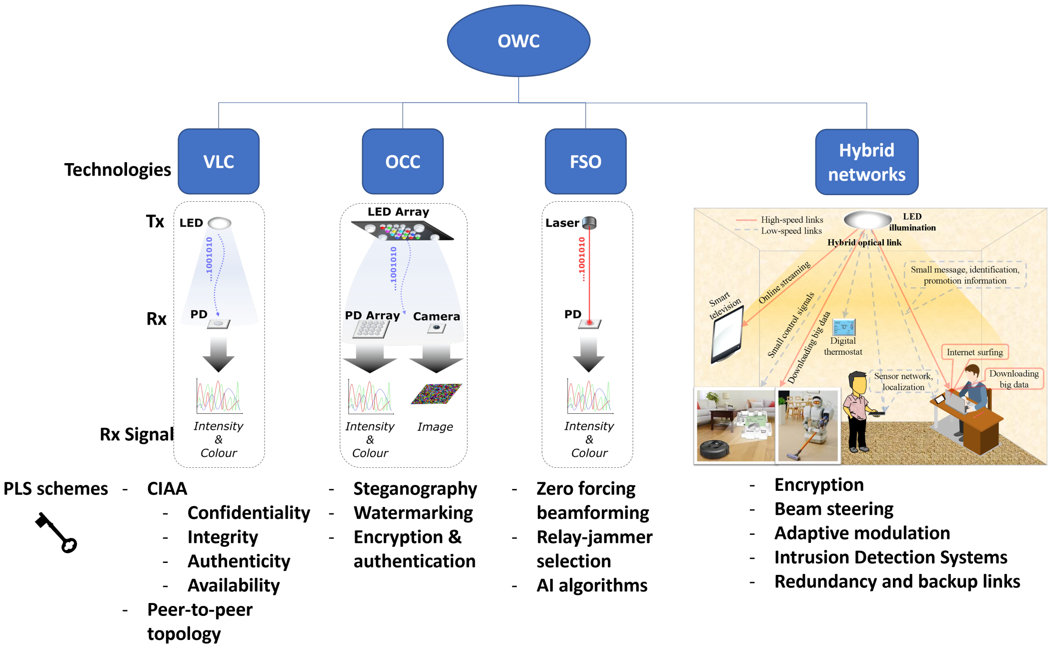


Figure 2.1: PLS schemes within OWC.

such as multi-antenna systems and cooperative nodes. In OWC systems, there are several PLS approaches to improve vulnerabilities as indicated in Fig. 2.1 for VLC, OCC, free space optics, and hybrid networks. The following sections bring a summary of state-of-the-art and future directions within OWC PLS.

2.3.2 State of the Art in OWC-based Physical Layer Security

Secret key generation

Access to a shared source of randomness can be exploited to generate secret keys between multiple agents by selectively choosing the samples from the shared source of randomness that provide an advantage with respect to potentially malicious agents [1]. Similarly, channels that present asymmetry between legitimate end agents and eavesdroppers (known as wiretap channels) can be also exploited for secret key agreements in a similar way, by distilling only those key segments in which the legitimate agents estimate that they had an advantage over the eavesdropper [1]. In [2], the authors report an implementation of a secret key generation scheme that uses deep learning to provide joint channel estimation and quantization, while the information reconciliation is based on the error correction approach.

PLS within VLC

In recent years, PLS in VLC has become an important area of research to improve privacy in wireless networks and work alongside encryption techniques used at higher network layers. The secrecy capacity of a VLC network is shown to be much higher than that of a WiFi network, mainly because of the unique properties of the THz band [3]. Since light cannot pass through walls, VLC networks can easily control the area's efficiency and security based on specific needs. Additionally, OWC networks do not experience small-scale fading, meaning the channel conditions are mostly determined by the positions of the transmit LEDs and receive PDs. This makes it possible to predict the channel response for the legitimate user

(Bob) if the optoelectronic characteristics of the LEDs, PDs, and Bob's 3D position are known to Eve. In the research, PLS in VLC is addressed using techniques like (i) beamforming, (ii) friendly jamming, (iii) mapping, and combinations of these methods.

Security vulnerabilities within VLC

Recently it has been demonstrated that VLC is prone to certain types of attacks, exploiting the fact that VLC is a wireless technology. For instance, by following the principle "what you see is what you get", which is very related to VLC, malicious nodes can intercept the visible light signals and create confidentiality issues or impersonation attacks. Authors in [4] demonstrated the feasibility of eavesdropping attacks, by showing data leakage. The interesting point to be considered and addressed in the context of VLC is that current solutions developed for RF-based systems are in general not directly applicable, due to the specific features of VLC. This makes it clear that new approaches, conceived by keeping in mind the key properties of VLC systems, need to be developed. It is with this principle in mind, that authors in [5] proposed and implemented LIBERO, an acronym of LIght Bias as an effective countermeasure against EavesdROpper attacks. The principle is simple, while effective. The idea is to implement a light-bias approach, to secure the communication system between a transmitter and a receiver. This light bias results in a signature for the end-to-end communication. Hence, the point is that with a predefined frequency the mapping between symbols and bits change over time and, even though the eavesdropper may detect the right symbol, it is not able to map it into the right bit sequence.

Moreover, in camera receiver-based VLC, i.e., OCC, steganography can be introduced as the technique of embedding secret information within optical signals, such as light or image data, transmitted via a camera system. This technique enables secure communication by concealing messages in a way that makes them undetectable to the human eye, yet recoverable by a receiver using specialized methods. It leverages the inherent properties of optical systems, such as modulation of light intensities or altering pixel values in images, to hide data without disrupting the normal functioning of the communication system. This approach finds applications in secure data transmission, surveillance, and watermarking. Recent studies have explored various methods to enhance the robustness and efficiency of optical steganography in the context of camera-based communications [6].

Visible light positioning and localization

OWC is a technology that has the property of spatial confinement as an inherent property which leads to two attractive application domains: i) deployment of dense, parallel communication channels that are nicely separated in space, which is especially attractive in IoT environments where a large number of nodes are involved, and ii) indoor positioning. This property makes eavesdropping in adjacent rooms becomes close to impossible. Moreover, by logging the initial position of the legitimate user and embedding this information in the light communication data, eavesdropping can be suppressed. Integrating Visible light positioning (VLP) and PLS allows precise user location tracking while protecting data from interception and attacks. This integration allows for secure and accurate positioning in environments like indoor settings, where both the location of devices and the confidentiality of the data are critical, especially for future communication systems like 6G. In specific cases, narrow optical beams' steering can be further used to minimize the area for eavesdropping. A novel single LED-based uplink VLP scheme with singular value decomposition beamforming was investigated in [7], to overcome challenges such as (i) raw data processing for camera-based positioning, (ii) extra image and video processing tools, and (iii) addressing privacy concerns related to cameras in public places.

PLS and Security vulnerabilities within Free Space Optics

Point-to-point FSO links provide an inherently asymmetrical channel if the elements are properly aligned and reflections are accounted for. In the presence of an eavesdropper, the system can always be assessed from the PLS point of view using a wiretap channel model. A theoretical analysis of the security of an FSO link in terms of the secrecy capacity is provided in [7]. In a point-to-multipoint link, e.g., Li-Fi, the optics of the transmitter provide a large coverage area, thus making the channels symmetric for a legitimate user and an eavesdropper. Using custom and tunable optical elements at the transmitter device, beamforming techniques can be implemented to create the asymmetry required for PLS. An analysis of the PLS of OWC systems is provided in [8] in terms of the secrecy outage probabilities and secrecy capacities under different attack scenarios. In [9], the security of simultaneous lightwave information and power transfer is analyzed.

Potential side-channels available in an FSO link are: (i) hidden receiver nodes behind or close to the legitimate receiver node; (ii) reflections coming from the receiver optics or other reflective surfaces (iii) stray signals coming from scattered or refracted beams, either close to the transmitter of to the receiver. In all of these cases, the attack model at the physical layer is modeled using Wyner's wiretap channel or, in the case of exploiting the turbulent nature of the atmosphere, a time-varying version of the same model.

Hybrid networks, IoT including OWC

In environments with a high density of users, such as stadiums, airports, shopping malls, etc., multiple RF access points can lead to interference and degrade system performance. Integrating OWC with RF in hybrid networks mitigates this issue by utilizing existing lighting infrastructure for data transmission. The hybrid OWC/RF networks enhance data rates, minimize RF interference, and offer a cost-effective solution, making them well-suited for high-density user scenarios. However, as hybrid OWC/RF networks combine both optical and radio frequency technologies, addressing security concerns is crucial to prevent potential vulnerabilities arising from the open and broadcasting nature of OWC and the susceptibility of RF to eavesdropping and interference. Moreover, for all-optical hybrid wireless communication systems, zero-forcing beamforming approach and a minimum power allocation algorithm was provided to overcome the physical layer security challenges [10].

The PLS aspects of hybrid OWC/RF communication systems have been widely analyzed in recent literature, primarily in terms of secrecy capacity and secure outage probability performance [11]. Novel PLS algorithms based on zero-forcing beamforming techniques have been proposed for hybrid VLC/RF in order to mitigate eavesdropping on both RF and VLC networks [12]. Additionally, energy harvesting [13] and intelligent reflecting surface (IRS) [14] have been introduced into hybrid VLC/RF to improve PLS characteristics.

2.3.3 Conclusion - Challenges and Future Work

PLS in OWC faces several challenges such as (i) vulnerability to eavesdropping due to the broadcast nature of light signals, especially in open or shared spaces and (ii) channel variations caused by obstacles, scattering by particles (fog, dust, etc) within the area optical beams have to pass, reflections, and user mobility can affect secure transmission. Implementing advanced encryption and beamforming techniques requires complex signal processing and additional computational resources. Future research should focus on developing adaptive PLS schemes that can dynamically adjust to changing environments. Machine learning and AI can enhance security by predicting and mitigating potential threats. Moreover, integrating PLS with VLP and multiple-input multiple-output techniques can further strengthen data protection in 6G and beyond communication systems.

2.3.4 References

[1] Djordjevic, Ivan B., "Continuous Variable (CV)-QKD," in *Physical-Layer Security, Quantum Key Distribution, and Post-Quantum Cryptography*, Cham, Switzerland: Springer Nature Switzerland, pp. 425–473, 2025. Available: https://doi.org/10.1007/978-3-031-88372-9_9

[2] Mahalal, Elmahedi; Ismail, Muhammad; Wu, Zi-Yang; Fouda, Mostafa M.; Fadlullah, Zubair Md; and Kato, Nei, "Robust Deep Learning-Based Secret Key Generation in Dynamic LiFi Networks Against Concept Drift," in *Proceedings of the 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*, pp. 899–904, IEEE, Jan. 2024. Available: https://doi.org/10.1109/CCNC51664.2024.10454770

[3] Blinowski, Grzegorz J., "The Feasibility of Launching Rogue Transmitter Attacks in Indoor Visible Light Communication Networks," *Wireless Personal Communications*, vol. 97, no. 4, pp. 5325–5343, Dec. 2017. Available: https://doi.org/10.1007/s11277-017-4781-3

[4] Loscri, Valeria; and Biagi, Mauro, "LIBERO: LIght Bias as Effective Countermeasure Against Eavesdropper Attacks," *IEEE Transactions on Communications*, vol. 72, no. 12, pp. 7882–7893, Dec. 2024. Available: https://doi.org/10.1109/TCOMM.2024.3420703

[5] Yin, He; Zhou, Xi; Xin, Nian; Hong, Jiaying; Li, Qin; and Zhang, Xiao, "Optical Steganography with Sign-Based Keys and Video as Vessel Medium," *Optics Communications*, vol. 526, article 128829, Jan. 2023. Available: https://doi.org/10.1016/j.optcom.2022.128829

[6] Zia-Ul-Mustafa, Rida; Le Minh, Hoa; Ghassemlooy, Zabih; Zvánovec, Stanislav; Isam Younus, Othman; Li, Xicong; and Pham, Anh T., "A Novel Uplink Positioning and SVD-Based Physical Layer Security Scheme for VLC Systems," *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 5, pp. 1706–1720, May 2025. Available: https://doi.org/10.1109/JSAC.2025.3543491

[7] Zia-Ul-Mustafa, Rida; Le Minh, Hoa; Ghassemlooy, Zabih; Zvánovec, Stanislav; Isam Younus, Othman; Li, Xicong; and Pham, Anh T., "A Novel Uplink Positioning and SVD-Based Physical Layer Security Scheme for VLC Systems," *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 5, pp. 1706–1720, May 2025. Available: https://doi.org/10.1109/JSAC.2025.3543491

[8] Abumarshoud, Hanaa; Chen, Cheng; Islim, Mohamed Sufyan; and Haas, Harald, "Optical Wireless Communications for Cyber-Secure Ubiquitous Wireless Networks," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 476, no. 2242, article 20200162, Oct. 2020. Available: https://doi.org/10.1098/rspa.2020.0162

[9] Liu, Xiaodong; Wang, Yuhao; Zhou, Fuhui; Ma, Shuai; Hu, Rose Qingyang; and Ng, Derrick Wing Kwan, "Beamforming Design for Secure MISO Visible Light Communication Networks with SLIPT," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7795–7809, Dec. 2020. Available: https://doi.org/10.1109/TCOMM.2020.3019818

[10] Chowdhury, Mostafa Zaman; Hasan, Moh Khalid; Shahjalal, Md; Hossan, Md Tanvir; and Jang, Yeong Min, "Optical Wireless Hybrid Networks: Trends, Opportunities, Challenges, and Research Directions," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 930–966, 2020. Available: https://doi.org/10.1109/COMST.2020.2966855

[11] Liao, Zhaohui; Yang, Liang; Chen, Jianchao; Yang, Hong-Chuan; and Alouini, Mohamed-Slim, "Physical Layer Security for Dual-Hop VLC/RF Communication Systems," *IEEE Communications Letters*, vol. 22, no. 12, pp. 2603–2606, Dec. 2018. Available: https://doi.org/10.1109/LCOMM.2018.2873725

[12] Al-Khori, Jaber; Nauryzbayev, Galymzhan; Abdallah, Mohamed M.; and Hamdi, Mounir, "Secrecy Performance of Decode-and-Forward Based Hybrid RF/VLC Relaying Systems," *IEEE Access*, vol. 7, pp. 10844–10856, 2019. Available: https://doi.org/10.1109/ACCESS.2019.2891678

[13] Pan, Gaofeng; Ye, Jia; and Ding, Zhiguo, "Secure Hybrid VLC-RF Systems with Light Energy Harvesting," *IEEE Transactions on Communications*, vol. 65, no. 10, pp. 4348–4359, Oct. 2017. Available: https://doi.org/10.1109/TCOMM.2017.2709314

[14] Zhang, Wei; Zhao, Xiang; Zhao, Yuqing; and Sun, Jinyong, "On Security Performance Analysis of IRS-Aided VLC/RF Hybrid System," *Physical Communication*, vol. 61, article 102176, Dec. 2023. Available: https://doi.org/10.1016/j.phycom.2023.102176

2.4 Security Aspects of Bidirectional Hybrid Millimeter-Wave Links based on Directly Modulated Lasers

Authors: M. Botella-Campos[1], J. Bohata[2], J. Romero-Huedo[1], S. Zvanovec[2], J. Mora[1], B. Ortega[1]

[1] Instituto de Telecomunicaciones y Aplicaciones Multimedia (ITEAM), Universitat Politècnica de València, Valencia, Spain

[2] Department of Electromagnetic Field, Czech Technical University in Prague, Faculty of Electrical Engineering, Prague, Czech Republic

2.4.1 Introduction

As 5G mobile networks continue to be deployed worldwide, both academia and industry are focusing on the future development of 6G systems. The evolution of mobile communications is driven not only by increasing mobile traffic and subscriptions but also by emerging applications such as immersive reality, interactive holography, and multisensory internet experiences. To address these demands, 6G technological enablers can be categorized into key areas: new spectrum, air interface, networking, architecture, and a fundamental paradigm shift [1].

Since enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC) will compete within the Internet of Everything (IoE), the 3G Partnership Project (3GPP) has introduced Releases 16 and 17 to enhance 5G New Radio (5G-NR) [2]. These advancements aim to meet the rigorous performance demands of future 6G systems, introducing three new scenarios: ubiquitous MBB (uMBB) for seamless global connectivity, ultra-reliable low-latency broadband communication (ULBC) for high-throughput, low-latency applications, and massive URLLC (mULC), which merges mMTC and URLLC capabilities. Thus, future networks must meet stringent Key Performance Indicators (KPIs), including 1 Gbps user experience, peak data rates of 1 Tbps, latency as low as 10 μs, and connectivity density of 10 ^{7} devices per km ^{2} [3].

Millimeter-wave (mmW) technology, first introduced in 5G new radio (NR), remains crucial for 6G networks due to its vast bandwidth, new air interface, and the adoption of open-radio access networks (O-RAN). Analog radio-over-fiber (RoF) solutions ensure seamless mmW signal transport, supporting low-latency and high-bandwidth links, particularly in small-cell environments where optical infrastructure costs are critical. The centralized network architecture consolidates baseband units (BBUs) at the central office (CO), optimizing resource allocation and extending coverage through remote radio heads (RRHs). Microwave-photonics-based techniques offer low phase noise, frequency tunability, and reduced reliance on electronic components, enabling cost-effective, low-latency, and high-bandwidth solutions. Among them, optical frequency multiplication using a Mach-Zehnder modulator (MZM) biased for carrier suppression (CS) has proven to be an effective approach for generating mmW signals. The use of a directly modulated laser (DML) with a CS-MZM is considered a promising solution for energy-efficient network deployment, as it provides a significant electrical gain in remotely generated mmW signals due to the combined effects of DML chirp and fiber dispersion [4].

Moreover, wireless mmW communications provide enhanced security through their inherent physical properties and advanced signal processing techniques. Due to their high frequency, mmW signals experience rapid attenuation and limited propagation, reducing the risk of eavesdropping over long distances. Their narrow beamwidth enables highly directional transmissions, making it difficult for attackers to intercept signals without being physically aligned with the communication path. Additionally, mmW systems leverage beamforming and spatial diversity, further minimizing unauthorized access. The high absorption of mmW signals by obstacles like walls enhances physical security by preventing signal leakage outside intended areas, which can even be reduced by operating at specific frequency regions of oxygen resonances (50-70 GHz, 118 GHz) or water vapour resonances (22.2 GHz). Advanced encryption and frequency agility in mmW networks add extra layers of protection against jamming and spoofing attacks. Moreover, the integration of secure beamforming and adaptive modulation techniques allows real-time adjustments to mitigate interference and potential threats. The complexity of mmW hardware and signal processing also raises the barrier for adversaries attempting unauthorized access. Emerging technologies, such as intelligent reflecting surfaces (IRS), further improve security by dynamically controlling signal paths.

2.4.2 State of the Art

Full-duplex RF signal transmission is essential for real-world applications, but many solutions require a laser source at the RRH for uplink (UL) transmission, increasing complexity and cost. This challenge can be addressed through carrier aggregation or wavelength reuse in centralized networks. To mitigate high-frequency distortion and bandwidth demands in the UL, mmW networks often transmit signals at lower frequencies. Various techniques, such as four-wave mixing and drop-filter demultiplexing, have been explored for remote local oscillator (LO) delivery. Recently, directly modulated lasers (DMLs) have been used in bidirectional photonic fronthaul links with external modulation for mmW signal generation, offering a simple, cost-effective, and high-performance solution for future 6G networks [5].

Among the centralized approaches for implementing photonics-assisted bidirectional mmW networks, a recently proposed system utilizes phase modulation (PM) technique [6] at both the CO and RRH to enable optical frequency up-conversion of the downlink (DL) signal and optical modulation of the down-converted UL signal, as shown in Figure 2.2. Compared to intensity modulators, PM offers key advantages, including stable DL up-conversion and centralized UL optical carrier generation from the CO with low modulator losses. Additionally, the system implements frequency down-conversion of the 40 GHz UL signal using an optically generated LO at the RRH, while the tunable laser source (TLS) for UL data transmission remains at the CO, simplifying remote site equipment. An optical waveshaper functions as a programmable filter, supporting DL frequency up-conversion, LO generation, and UL optical carrier provision. Insets in Figure 2.2 allow us to follow the signal generation basics showing the spectrum of CS modulated signal after PM1 and spectra at photodiodes in the RRH where each optical carrier is selected for mmW signal generation at photodiode PD1 and LO signal generation at PD2.

The downconverted UL mmW signal is employed to phase modulate the TLS optical carrier at RRH and the signal is recovered after UL transmission and demultiplexing at the CO. The details of the DL and UL recovered electrical spectra and signal constellation are also shown as insets in Figure 2.2.

Figure 2.3 shows a photograph of the experimental setup. Validation measurements over a 10 km standard single mode fiber (SSMF) and 1.1 m long radio wireless links tested 64-quadrature amplitude modulation (64-QAM) at 41 GHz for the DL and quadrature phase shift keying (QPSK) at 40 GHz for the UL. The system successfully transmitted up to 200 MHz bandwidth for both modulation formats while maintaining error vector magnitude (EVM) well below the specified threshold, as shown in Figure 2.3 (b) and (c), for different received electrical power (ReP) in the DL and UL, respectively [6]. The DL in the full configuration, which outperforms the system without antennas, shows consistent performance trends, regardless of whether the uplink (UL) is active and requires a minimal ReP of -43 dBm for

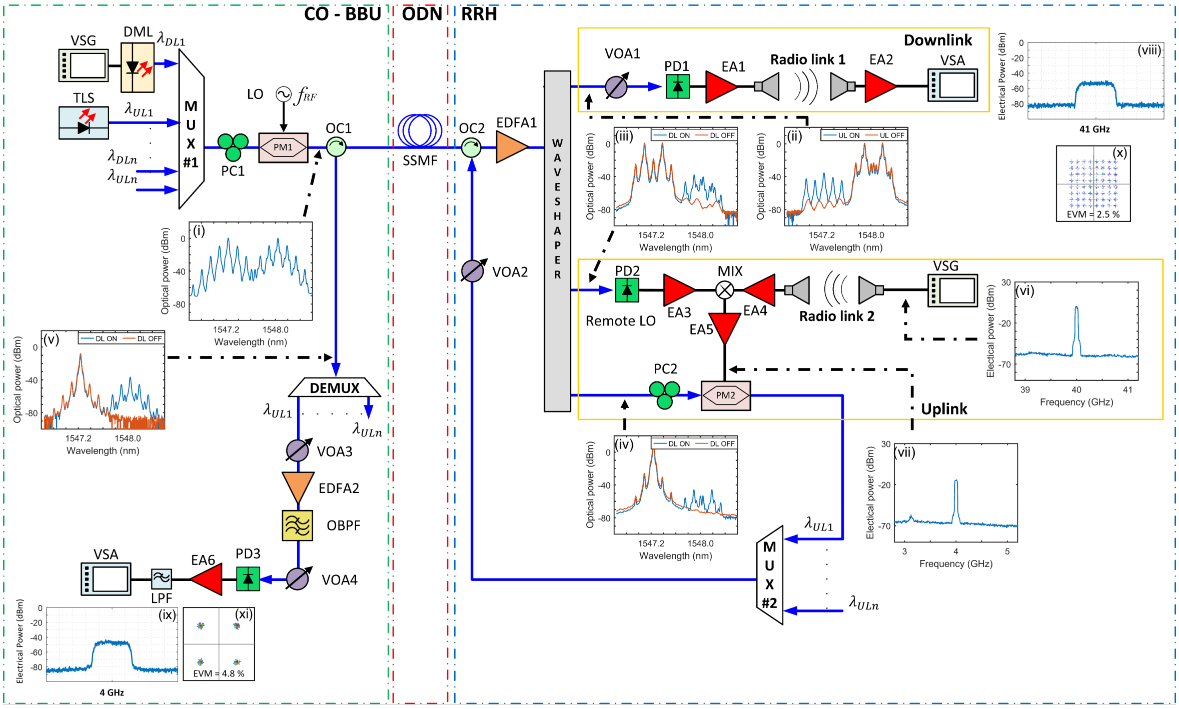


Figure 2.2: Schematic of the experimental setup of the PM-based full-duplex mmW fronthaul link with LO delivery for UL down-conversion. Insets show the optical (i-v) and electrical (vi-ix) spectra at different points. Insets (x) and (xi) shows received constellations at DL and UL, respectively [6].

successful transmission. In the UL, the Radio + OB2B configuration achieves the highest signal-to-noise ratio (SNR) in contrast with the Full Link scenario with the DL ON.

In contrast, in the UL characterization, the SNR diminishes, i.e., EVM increases, for the Full Link scenario. Note that a bidirectional transmission system with optical carrier reuse, reflections and Rayleigh backscattering leads to undesirable degradation since the phase noise of the optical carrier could be converted to intensity noise through the abovementioned interferences. However, measured EVM values remain below the threshold and successful transmission is demonstrated across the range of ReP values. Moreover, further tests in full-duplex operation with 5G-NR orthogonal frequency-division multiplexing (OFDM) signals at 100 MHz bandwidth showed EVM values as low as 5.2 % for the DL and 6.9 % for the UL, demonstrating the effectiveness and robustness of the proposed solution.

The security in mmW wireless communications leverages advanced physical-layer security techniques, beamforming, and cryptographic mechanisms to mitigate threats. Emerging technologies, such as IRS and reconfigurable metasurfaces, enhance security by dynamically shaping signal propagation and minimizing interception opportunities [7]. Physical-layer security techniques, including artificial noise injection and secrecy coding, further strengthen protection against passive and active attacks. In Simultaneous Wireless Information and Power Transfer (SWIPT), IRS optimizes beamforming to ensure secure energy and data transmission, reducing signal leakage. Integrating IRS with Non-Orthogonal Multiple Access (NOMA) in Integrated Sensing and Communication (ISAC) systems has improved the sum secrecy rate, making secure multi-user communication more efficient. Active RIS further enhances secrecy by amplifying desired signals while suppressing potential threats, particularly in high-frequency mmW and THz bands. In 6G networks, RIS creates adaptive and recon-

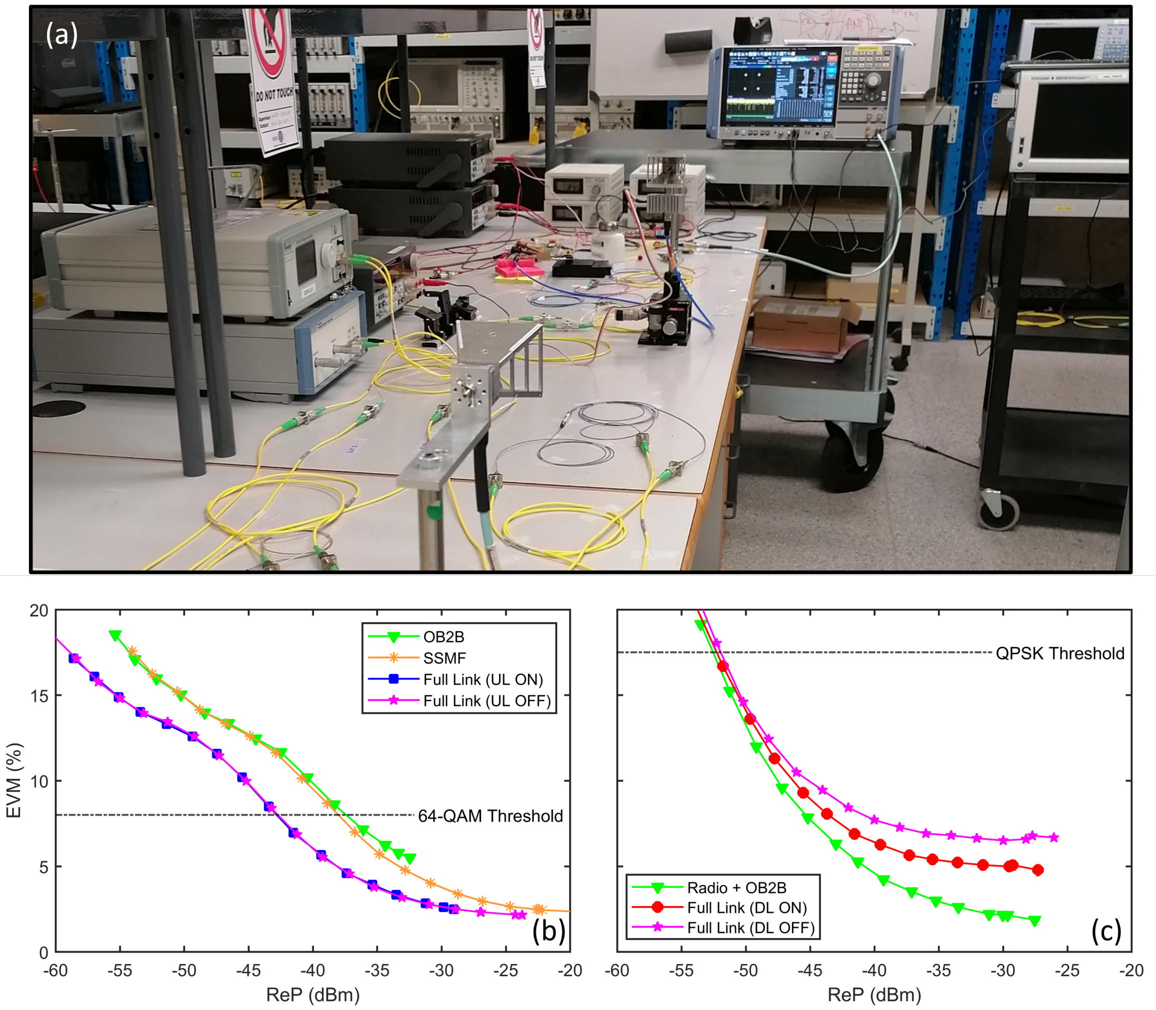


Figure 2.3: (a) Photograph of the experimental setup, EVM measurements at mmW frequency of 41 GHz versus ReP under OB2B and Full Link (UL ON), and Full Link (UL OFF) operation for DL (b) and UL (c). The former one also shows EVM after fiber propagation (SSMF).

figurable channels that prevent unauthorized access, offering privacy-preserving solutions by obscuring signal paths and user locations.

Emerging security approaches, such as packet erasures and directional beamforming, show great promise for secure key exchange in 6G and autonomous vehicle networks. To counteract computational attacks in mmW networks, integrating robust encryption protocols, including quantum-resistant cryptography, is essential. Additionally, Continuous Variable Quantum Key Distribution (CVQKD) using mmW and THz frequencies, supported by cryogenic technology, offers a potential breakthrough for secure wireless communications and the future quantum internet [8].

2.4.3 Challenges and Future Work

Photonics-assisted RoF technology is set to play a crucial role in 5G and beyond, offering high-capacity, low-latency, and cost-effective network solutions. However, challenges remain in optimizing remote components, improving mmW transmission, and refining optical modulation techniques for widespread deployment. While mmW communications enhance cybersecurity by reducing eavesdropping risks and enabling secure high-speed data transfer, they are vulnerable to physical disruptions, require complex hardware, and demand high infrastructure investments. The future of secure RoF networks lies in integrating AI-driven security, adaptive beamforming, and advanced encryption, ensuring stronger resilience against cyber threats. Researchers are actively working to address vulnerabilities such as jamming, beam tracking attacks, and device authentication risks by developing hybrid security models that blend physical-layer defenses with cryptographic techniques, paving the way for robust, next-generation wireless networks.

2.4.4 References

[1] Tataria, Harsh; Shafi, Mansoor; Dohler, Mischa; and Sun, Shu, “Six Critical Challenges for 6G Wireless Systems: A Summary and Some Solutions,” *IEEE Vehicular Technology Magazine*, vol. 17, no. 1, pp. 16–26, Mar. 2022. Available: https://doi.org/10.1109/MVT.2021.3136506

[2] Dao, Nhu-Ngoc; Tu, Ngo Hoang; Hoang, Trong-Dai; Nguyen, Tri-Hai; Nguyen, Luong Vuong; Lee, Kyungchun; Park, Laihyuk; Na, Woongsoo; and Cho, Sungrae, “A Review on New Technologies in 3GPP Standards for 5G Access and Beyond,” *Computer Networks*, vol. 245, article 110370, May 2024. Available: https://doi.org/10.1016/j.comnet.2024.110370

[3] Jiang, Wei; Han, Bin; Habibi, Mohammad Asif; and Schotten, Hans Dieter, “The Road Towards 6G: A Comprehensive Survey,” *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021. Available: https://doi.org/10.1109/OJCOMS.2021.3057679

[4] Botella-Campos, M.; Mora, J.; and Ortega, B., “MMW Signal Gain in DML-Based Microwave Photonic Links Under Large Signal Regime,” *Journal of Lightwave Technology*, vol. 41, no. 18, pp. 5991–5999, Sep. 2023. Available: https://doi.org/10.1109/JLT.2023.3277701

[5] Vallejo, Luis; Bohata, J.; Mora, J.; Zvánovec, S.; and Ortega, B., “Remote mmW Photonic Local Oscillator Delivery for Uplink Down-Conversion in DML-Based Optical Hybrid C-RAN Fronthaul,” *Journal of Optical Communications and Networking*, vol. 15, no. 6, pp. 357–366, Jun. 2023. Available: https://doi.org/10.1364/JOCN.482085

[6] Bohata, J.; Botella-Campos, M.; Mora, J.; Ortega, B.; and Zvánovec, S., “Centralized Full-Duplex Seamless Photonic mmW Fronthaul Link Based on Phase Modulation,” *Journal of Optical Communications and Networking*, vol. 16, no. 6, pp. 670–680, Jun. 2024. Available: https://doi.org/10.1364/JOCN.514977

[7] Khan, Adil; Mohsan, Syed Agha Hassnain; Elfikky, Abdelrahman; Boghdady, Ayman I.; Ahmad, Shabeer; and Innab, Nisreen, “A Survey of Intelligent Reflecting Surfaces: Performance Analysis, Extensions, Potential Challenges, and Open Research Issues,” *Vehicular Communications*, vol. 51, article 100859, Feb. 2025. Available: https://doi.org/10.1016/j.vehcom.2024.100859

[8] Zhang, Mingqi; Pirandola, Stefano; and Delfanazari, Kaveh, “Millimeter-Waves to Terahertz SISO and MIMO Continuous Variable Quantum Key Distribution,” *IEEE Transactions on Quantum Engineering*, vol. 4, pp. 1–10, 2023. Available: https://doi.org/10.1109/TQE.2023.3266946

2.5 Quantum Secure Communications for Emerging Wireless Systems and 7G

Authors: Tuğrul Çavdar[1], Iraklis Symeonidis[2]

[1] Department of Computer Engineering, Karadeniz Technical University, Trabzon, Türkiye
[2] Department of Industrial Systems, Unit Smart Automation and Cyber Resilience, RISE, Sweden

2.5.1 Introduction

Every new generation of wireless mobile networks brings significant updates in terms of communication technologies, speed of operation, network parameters, operating frequencies, security, etc. The 7G (Seventh Generation) is the intelligent cellular technology that will succeed the 5G and 6G, by upgrading to a much higher frequency range, a higher capacity and a much lower delay in communication. 7G is expected to provide the capabilities of non-existent delay in communication, very high-speed connectivity, utilization of Artificial Intelligence based core networking solutions, virtual space environment with realistic sensations, internet cognition, better global coverage by using a satellite network, internet of everything, better remote access for diagnosis, learning, etc.

By introducing quantum technology, 7G will be able to speed up computing processes and enable use of quantum computing based cryptography as well as to enable distributed quantum computing, that is a new paradigm in computer sciences. With the advent of quantum computers, Shor's algorithm, a quantum algorithm that finds prime factors of an integer, and quantum search algorithms have compromised the RSA and ECC schemes that run on classical computers. This paved the way for research into post-quantum and quantum cryptography algorithms for use in 7G networks.

This paper provides an overview of Quantum Secure Communications techniques, and gives challenges and future directions for the potential areas of use in emerging wireless networks and 7G.

2.5.2 State of the Art: Secure Quantum Communication Protocols

The growing threat of cyberattacks has highlighted the vulnerability of classical cryptographic methods. In response to this challenge, quantum cryptography has emerged as a revolutionary solution promising unbreakable encryption. This section focuses on the secure quantum communication protocols.

Quantum Key Distribution

Quantum Key Distribution (QKD) enables two parties to produce and share a key via quantum channel by relying on fundamental laws of quantum mechanics that protects the data, such as, the no-cloning theorem makes impossible to create identical copies of a quantum state, which prevents attackers from copying the data. Additionally, if an attacker listens to the communications (i.e., measures the quantum states), the system will change in such a way that both sender and receiver parties will know it. The follow-up communication of the message encrypted with the key is still based on classical communications.

QKD Protocols are:

- BB84 (Charles H. **B**ennett, Gilles **B**rassard, 19**84**, HUP, DV)
- E91 (Artur **E**kert, 19**91**, QE, DV)

- B92 (Charles H. **B**ennett, 19**92**, HUP, DV)
- BBM92 (Charles H. **B**ennett , Gilles **B**rassard, N. David **M**ermin, 19**92**, QE)
- MSZ96 (Yi **M**u, Jessica **S**eberry, Yuliang **Z**heng, 19**96**)
- Six-state (Dagmar Bruss, 1998)
- SSP (Bechmann, H. Pasquinucci, Nicolas Gisin, 1999 , HUP)
- Silberhorn (Christine **Silberhorn**, 2001, QE, CV)
- DPS (**D**ifferential-**P**hase-**S**hift, Kyo Inoue, Edo Waks, Yoshihisa Yamamoto, 2002, QE, DV)
- GG02 (Frédéric **G**rosshans and Philippe **G**rangier, 20**02**, HUP, CV)
- Decoy state (HoiKwong Lo, Xiongfeng Ma, Kai Chen, 2003, HUP, DV)
- SARG04 (Valerio **S**carani, Antonio **A**cin, Gregoire **R**ibordy, Nicolas **G**isin, 20**04** , HUP, DV)
- COW (**C**oherent **O**ne-**W**ay, D. Stucki, N. Brunner, Nicolas Gisin, Valerio Scarani, H. Zbinden, 2005, QE, DV)
- KMB09 (Muhammad Mubashir **K**han, Michael **M**urphy and Almut **B**eige, 20**09** , HUP)
- S13 (Eduin Esteban Hernandez **S**erna, 20**13** , HUP)
- HD-QKD (Yun**H**ong Ding, **D**avide Bacco, 2017)

where the principles are *HUP: Heisenberg's Uncertainty Principle* and *QE: Quantum Entanglement* and where the types are *CV: Continuous Variable* and *DV: Discrete Variable.*

There is no single "best" QKD method. It depends on the application, performance requirements, and infrastructure. However, here is the strengths of the most commonly used QKD protocols:

- Most secure: Device-Independent QKD (DI-QKD)
- Best tradeoff between security and practicality: Measurement-Device-Independent QKD (MDI-QKD)
- Most widely used and easy to implement: BB84 Protocol. It is also the best for satellite-based QKD.
- Best for long distances: Twin-Field QKD (TF-QKD)

Quantum entanglement takes place when two or more particles interact in a manner such that the state of one cannot be described without referencing the state of the other(s), no matter the separating distance. This is to say that measuring one particle directly affects the state of the other entangled particle(s). Heisenberg's uncertainty principle asserts that it is impossible to know simultaneously certain pairs of physical properties, such as position and momentum, with unlimited precision. The more precisely one is known, the less precisely the other may be known.

In DV-QKD, single photons are sent through the channel, one at a time. The information can be encoded in the polarization of each photon. It is the best suited for long-haul networks. In CV-QKD, continuous beam of light is sent, as in most classical optical communications. The information can be encoded by modulating the amplitude and phase of the electromagnetic wave. It gives the highest performance in metro networks.

Quantum Secure Direct Communication

Quantum Secure Direct Communication (QSDC) provides direct and secure communications between two parties via quantum channels based on quantum mechanics without using a security key distribution [1].

Quantum Identity Authentication

Quantum Identity Authentication (QIA) guarantees that the recipient gets the qubits from the right sender, and the sender guarantees that the output of the quantum channel is read by the authentic receiver [2]. In QIA, Alice is an authenticator, Bob_1, Bob_2, $\cdot \cdot \cdot$, Bob_r are the certified users. Trent, which is the third party, helps Alice to simultaneously authenticate Bob_1, Bob_2, $\cdot \cdot \cdot$, Bob_r. The steps are as follows:

1. Trent sends secret keys to Alice and Bobs at the time of registration. Alice transmits qubits to the Bob_1, Bob_2, $\cdot \cdot \cdot$, Bob_r individually during preparation.
2. Bob_1, Bob_2, $\cdot \cdot \cdot$, Bob_r send the measured and operated particles to Alice, respectively.
3. Alice reports her computation results to Trent.
4. Trent compares results to see if the authentication is successful or not, and announces it to all users at the same time.

Quantum Secret Sharing

Secret Sharing shares a secret key among several users in a way that they are all given a share of the secret key without any one of them individually being able to figure out the whole secret key. The secret key can only be reconstructed upon cooperation of the nodes. Such as,

Alice's Key + Bob's Part = Charlie's Part

Quantum Secret Sharing (QSS) enables splitting and sharing the secret key among multiple nodes on quantum states instead of classical bits.

QSS Schemes are [3]:

- QSS sharing classical information
- QSS sharing quantum state
- QSS based on the single particle
- QSS based on the maximal entangled states
- QSS based on the non-maximal entangled states
- QSS realizing the (n-n) structure (data is divided into n pieces and easily reconstructed from these n pieces)
- QSS realizing the (t-n) structure (data is divided into n pieces in such a way that data is easily reconstructable from any t pieces; namely, any t out of n participants can recover the secret)
- QSS realizing the general access structure
- QSS demonstrating the verifiable QSS
- QSS demonstrating the dynamic QSS
- QSS demonstrating the error-correcting QSS

In QSS sharing quantum state, Alice, Bob, and Charlie all possess one particle of a GHZ triplet. They all randomly decide to measure their particle in the x or y direction. The steps as the following [4]:

1. Alice combines her qubit with her GHZ particle, and measures the pair in the Bell basis. She does not tell the result to Bob and Charlie.
2. Alice tells Bob and Charlie to measure their GHZ particles. For instance Bob measures, then tells the result to Charlie.

3. Alice sends her measurement result to Charlie. So, Charlie learns the amplitude of Alice's qubits, but knows nothing about its phase.
4. Bob's qubit gives Charlie the phase information, and allows him to reconstruct Alice's qubit.

Bob not only has to measure his particle but also transmit the outcome to Charlie so that Charlie can reconstruct Alice's state with the help of Bob.

Other Quantum Security Topics

- Quantum Permutation Pad (a quantum-safe symmetric cryptographic algorithm [5])
- Quantum Consensus (reaching agreement in quantum networks) [6]
- Quantum Coin Flipping (a protocol based on quantum mechanics to be used between two or more parties who do not trust each other) [7]

2.5.3 Challenges and Future Work

The quantum computers will enable work on optimization of emerging large-scale networks and potential solutions for enhancing energy and decision latency efficiency in 7G. Quantum computers also offer the advantages quantum information processing when it comes to feasible practical implementation in a variety of applications in 7G like, resource availability prediction, optimum network resource allocation, traffic monitoring, network control, network design, deployment and operation. Space roaming with the support of QKD over quantum satellite network will provide high datarate and enhanced security.

Since the network optimization issue is complex, quantum computing provides the ultimate solution in order to be capable of dealing with the network dynamics. The optimization issues are combinatorial in nature and would be significantly helped by the application of quantum algorithms. Beyond the gigantic speed up in computation from the parallelism in the operation (Google has reported a quantum computer capable of computing 10^8 times faster than the classical one).

The primary restriction on the application of quantum computing technology in future networks is the requirement to utilize centralized instead of distributed information processing rendering the effects of propagation delays. With the development of distributed quantum computing, these restrictions can be alleviated to a certain degree.

2.5.4 References

[1] Wang, Chonggang; and Rahman, Akbar, "Quantum-Enabled 6G Wireless Networks: Opportunities and Challenges," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 58–69, Feb. 2022. Available: https://doi.org/10.1109/MWC.006.00340

[2] Azahari, Nur Shahirah Binti; Harun, Nur Ziadah Binti; and Zukarnain, Zuriati Binti Ahmad, "Quantum Identity Authentication for Non-Entanglement Multiparty Communication: A Review, State of Art and Future Directions," *ICT Express*, Mar. 2023. Available: https://doi.org/10.1016/j.icte.2023.02.010

[3] Qin, Huawang; and Dai, Yuewei, "Efficient Quantum Secret Sharing," *Quantum Information Processing*, vol. 15, no. 5, pp. 2091–2100, May 2016. Available: https://doi.org/10.1007/s11128-016-1251-x

[4] Hillery, Mark; Bužek, Vladimír; and Berthiaume, André, "Quantum Secret Sharing," *Physical Review A*, vol. 59, no. 3, pp. 1829–1834, Mar. 1999. Available: https://doi.org/10.1103/PhysRevA.59.1829

[5] Kuang, Randy; and Barbeau, Michel, "Quantum Permutation Pad for Universal Quantum-Safe Cryptography," *Quantum Information Processing*, vol. 21, no. 6, Jun. 2022. Available: https://doi.org/10.1007/s11128-022-03557-y

[6] Marcozzi, Marco; and Mostarda, Leonardo, "Quantum Consensus: An Overview," arXiv preprint, Jan. 2021. Available: https://arxiv.org/abs/2101.04192

[7] Kitaev, Alexei, "Quantum Coin Flipping," Quantum Information Processing Workshop, Mathematical Sciences Research Institute, University of California, Berkeley, 2003. Available: https://cir.nii.ac.jp/crid/1571698600695039360

Chapter 3

Foundations of Trust and Future Network Integrity

3.1 Introduction

Chapter Editors: Basak Ozan Ozparlak [1], Roya Khanzadeh[2,3], Ana Ferreira [4]

[1] Ozyegin University, Faculty of Law, Istanbul, Turkiye
[2]Johannes Kepler University Linz, Institute for Communications Engineering and RF-Systems, Linz, Austria
[3]JKU LIT SAL IWS Lab, Linz, Austria
[4] RISE-Health, Department of Community Medicine, Information and Health Decision Sciences Faculty of Medicine, University of Porto, Portugal

This section lays the groundwork for building trustworthy and high-assurance wireless systems in the 6G era. As 6G networks evolve toward increasingly intelligent, complex, and large-scale, the demand for trustworthiness will intensify. Therefore, it is vital to explore how trust assessment frameworks, architectural design principles, and regulatory compliance shape performance and reliability in next-generation networks. This section aims to contribute to this goal by examining methodologies for assessing, measuring, and enhancing trustworthiness and its key characteristics, such as resilience and robustness, while also exploring the role of legal compliance in enabling privacy-preserving, secure, and resilient 6G connectivity. Furthermore, it extends the discussion of trust and integrity beyond the network infrastructure, addressing multi-vendor supply chain risks and the evolving regulatory landscape in Europe.

For 6G to be a trustworthy network, it will need to move beyond static trust assumptions and implement continuous, real-time trustworthiness assessment and adaptation, grounded in dynamic trust models with zero-trust principles, which requires a new architectural paradigm. Trust also needs to be calculated, propagated, and managed across all network layers, using context-aware metrics and decision processes that consider both observable behavior and broader security policies. Since 6G will be an AI-native network, ensuring the trustworthiness of AI itself becomes crucial, requiring transparency, explainability, and robustness against manipulation to uphold overall network trustworthiness. Resilience and robustness are among the main characteristics defined in a trustworthy network. 6G networks must demonstrate the ability to withstand, adapt to, and recover from failures, whether accidental, systemic, or adversarial. As critical infrastructure becomes increasingly interdependent, identifying and protecting vital nodes and links will be essential to maintain operational continuity and public trust.

Besides trust assessment and resilience measurement, enhancing trustworthiness in 6G requires active and continuous management through network management tasks. Integrating trust-aware mechanisms into network management tasks such as scheduling, resource allocation, and routing enables the network to dynamically adapt security, reliability, and privacy levels while maintaining network key performance indicators (KPIs). This approach ensures a balanced, robust, and trustworthy 6G system that responds effectively to changing network conditions and user demands.

On the other hand, a dedication to Privacy by Design (PbD) principles is necessary to ensure and maintain trust in 6G networks, which are characterized by massive data flow and AI-driven services. Putting PbD principles into practice could help maintain and ensure trust. By directly integrating privacy protections into system architectures, PbD principles proactively reduce such risks. However, effective implementation depends on harmonizing fragmented regulatory compliance.

Finally, trust is not confined to the infrastructure, but also extends beyond it to the multi-vendor supply chain ecosystem. Consider supply chain security as an example: since 6G networks will be heterogeneous and built from components sourced globally across multiple vendors, the vulnerability of all network elements (including software, hardware, and third-party services) will increase. To address these risks, end-to-end visibility, robust risk management, and compliance with emerging regulatory frameworks such as the Cyber Resilience Act (CRA) and the EU's NIS2 Directive are essential.

Together, these insights form the foundation of a trustworthy, adaptive, and regulation-aware 6G architecture.

3.2 Trustworthy 6G Networks: Metrics, Methods, and Future Directions

Authors: Oleksandra Yeremenko[1], Roya Khanzadeh[2,3], David Jelenc[4]

[1]Kharkiv National University of Radio Electronics, V.V. Popovskyy Department of Infocommunication Engineering, Kharkiv, Ukraine
[2]Johannes Kepler University Linz, Institute for Communications Engineering and RF-Systems, Linz, Austria
[3]JKU LIT SAL IWS Lab, Linz, Austria
[4]University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia

3.2.1 Introduction

As the sixth-generation technology for wireless communication (6G) networks changes the modern communications concept, their trustworthiness becomes essential. Indeed, 6G brings together intelligent, heterogeneous networks and breaks the boundaries between the physical and digital worlds. In the 6G ecosystem, when billions of interconnected devices interact with new human-machine interfaces, taking into account the multi-vendor supply chain and the use of various software types, any security breach can have real-world consequences [1]. In addition, using artificial intelligence (AI) in 6G creates new vulnerabilities, making trustworthy AI essential to ensure secure and sustainable operations. 6G trustworthiness should become a fundamental layer of network architecture, ensuring reliability, security, privacy, and availability to address these issues. Built on trust, intelligence, and heterogeneity, 6G requires a paradigm shift in trustworthiness, which requires innovative solutions such as intelligent trustworthiness management, blockchain for secure transactions, AI-based security protocols, and strong authentication mechanisms. The successful development of trustworthy 6G depends on bridging research gaps, designing adaptive trustworthiness architectures and models, and establishing standardized trustworthiness metrics, assessment methods, and frameworks to ensure its integrity and performance.

However, current approaches to trustworthiness remain fragmented as they combine elements of security, cyber resilience, and trust management of network systems. The lack of unified approaches and clear definitions of trust, trustworthiness, security, and resilience creates methodological and practical challenges to build a trustworthy 6G network. Therefore, a systematic analysis of existing disciplines, their comparison and the development of common trustworthiness metrics are necessary to create transparent and effective trust mechanisms in the 6G architecture.

Trustworthiness is usually defined as the degree to which the entity can meet critical requirements or perform according to the designed behavior under any set of conditions [2]. From the communication system point of view, trustworthiness is described by its characteristics, including, but not limited to, reliability, security, robustness, resilience, availability, integrity, safety, privacy, confidentiality, and also, benevolence, functional or non-functional capacities, and intervenability [3].

On the other hand, trust is often understood as the extent to which one entity is willing to depend on another in a given situation [4]. While it is typically derived from trustworthiness, the derivation may include other factors, even subjective ones such as the utility of the interaction in question or the entity's risk tolerance.

To build up a trustworthy 6G network, we need a multi-layer hierarchical trustworthiness architecture (see Figure 3.1). At the innermost layer of this architecture is the user equip-

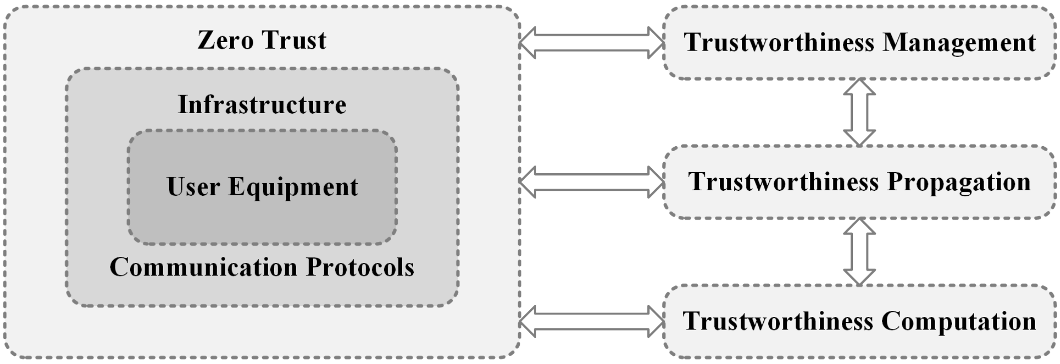

Zero Trust
Infrastructure
User Equipment
Communication Protocols
Trustworthiness Management
Trustworthiness Propagation
Trustworthiness Computation

Figure 3.1: Conceptual Framework for Multi-layer Hierarchical Trustworthiness Architecture.

ment, which comprises different mechanisms to ensure data trustworthiness in the physical devices and measure the degree of trustworthiness of end users' devices. At the inner layer of the hierarchy is the infrastructure, which ensures trustworthiness at the architectural level and communication protocols and links. The outermost layer provides fundamental trust and originates from three aspects: trusted foundation, trusted platform, and trusted hardware, ensuring the technologies involved in 6G are inherently designed to be trustworthy in order to preserve network trustworthiness [5]. The zero trust architecture is one promising approach that can be deployed in this hierarchy model, where trustworthiness assessment and computation continuously and dynamically evaluate users and devices. The rationale behind this approach is that users' behavior is not static but changes over time, influenced by various factors, such as network conditions and communication links, user activities and movement, and application usage [6].

A trustworthiness model must be developed at each layer of the above-described architecture with three main building blocks: trustworthiness computation, propagation, and management. A trustworthiness computation component makes an assessment based on predefined metrics, and it usually consists of techniques that aggregate observations to obtain a single trustworthiness score [7]. Trustworthiness propagation handles the dissemination of trustworthiness-related information of the involved entities in the network in a distributed, semi-distributed, or centralized fashion [8]. Finally, trustworthiness management is the core of the model, wherein the final decision on whether a user, communication link, node, or component is trusted or not is made, and measures and countermeasures for enhancing the overall network trustworthiness are optimized. While the decision to trust or not can be as simple as thresholding the trustworthiness, it is often more involved in incorporating contextual information, security policies, as well as subjective factors such as risk tolerance. Figure 3.1 shows a conceptual framework for the hierarchy architecture with its interactions with trustworthiness main building blocks.

3.2.2 State of the Art

6G Trustworthiness Metrics

Considering 6G networks' vulnerabilities and requirements, the challenge arises in identifying appropriate metrics to evaluate 6G network trustworthiness. One relevant approach is to establish appropriate metrics, which first requires defining potential threats and vulnerabilities affecting different 6G network components, including nodes (base stations, user devices, core network elements), links (wireless and wired transmission paths), and the entire system (network applications, AI-driven optimizations, API vulnerabilities, software risks, and protocol weaknesses). Once these threats are identified, the corresponding quantitative and qualitative metrics can be formulated to assess security, resilience, and overall 6G trustworthiness [9, 10].

Furthermore, metrics need to be categorized. These metrics can be primitive metrics obtained from raw data measurements or numerous variants of derived metrics based on possible aggregation functions related to key aspects of network trustworthiness (e.g., security, resilience, trust). At the same time, the modeling and analysis-based measurements stage is essential when proposing derived metrics. These measurements aim to quantify the performance attributes of the system or its components and explore the application of various analytical methods, simulation, and emulation tools.

Given the introduction of native AI, which characterizes 6G networks, identifying appropriate AI trustworthiness metrics is also essential. Moreover, it should be taken into account that according to the European Commission's Ethics guidelines, trustworthy AI should be lawful in respect of regulations, ethical in principles and values, and robust technically and socially. Among the trustworthiness metrics that can be evaluated for AI-based systems for 6G, we can highlight the following [11]. In data quality assessment, the metrics can be distinguished regarding data completeness, correctness, diversity, and representativeness. The AI operability metrics must prove a safe and reliable system function. While AI dependability metrics ensure critical AI systems deliver justifiably trusted service and meet functional safety requirements. Robustness metrics assessment and monitoring are also mandatory due to the demands for proper system functioning in adverse conditions and avoiding safety and security risks. Besides, the specific AI systems' explainability metrics aim to provide an interface between humans and AI. Finally, human-centered quality and human oversight metrics are connected with privacy and respect for fundamental human rights.

6G Trustworthiness Assessment Methods

Trustworthiness can be assessed in multiple ways. Here, we classify assessment methods along three dimensions: static or dynamic, local or global, and ad-hoc or by design. The first dimension distinguishes between static assessments, which are performed once and remain unchanged, and dynamic assessments, which are updated continuously [5]. For instance, computer hardware, which rarely changes during operation, can be evaluated statically. A suitable method for such an evaluation is the Common Criteria for Information Technology Security Evaluation, an international standard that assesses the reliability of a device based primarily on its security functions. In contrast, the trustworthiness of users, devices, and applications varies over time, requiring continuous evaluation. Trust models address this need by collecting behavioral data and estimating trustworthiness through data aggregation, heuristic-based computations, machine-learning algorithms, or sophisticated reasoning techniques [15].

Another key distinction in trustworthiness assessment concerns the data's origin, which can be local or global. Local assessments rely solely on data collected by a specific application or component, whereas global assessments incorporate data from multiple sources. For example, a base station that evaluates connected devices using only locally gathered data performs a local assessment. If the assessment were global, the base station would also consider data from other base stations. While this broader approach generally improves accuracy, it introduces new vulnerabilities: a base station must trust the reliability of other base stations, as incorrect or malicious data could corrupt the assessment. Calculating a device's global trustworthiness aligns with reputation models, which are designed to address such challenges [10, 8].

A third approach to trustworthiness assessment, advocated in [16], follows the principle of trustworthiness-by-design. This principle asserts that trustworthiness should be an integral

Table 3.1: Recent trustworthiness assessment frameworks overview

	[12]	[13]	[9]	[10]	[2]	[14]
Characteristics[1]	Com, Use, Fnc, Rob, Neu, Exp	Rel, Sec, Res, Saf, Prv	Rel, Sec, Res, Saf, Prv	Sec	Rel, Avl, Saf, Cnf, Int, Rob, Mnt, Adp, Use, Tim, Eff, Rct, Pro	Sec, Prv, Res
Metrics[2]	✗	✗	physical layer parameters	✗	✗	✗
Assessment level	trustworthiness model	4-layer[3]	node, link, system	4-layer	4-layer	network elements, parties in supply chain, organization in industry
Computation method[4]	✗	✗	dynamic, local, reasoning	✗	✗	✗
Grading[5]	binary, coarse-grained, fine-grained, semantic	coarse-grained, fine-grained	binary, fine-grained	✗	coarse-grained	✗
AI-trustworthiness	✓	✗	✗	✓	✗	✓
Propagation	✗	✗	✗	✓	✗	✓
Application	telecommunication networks	IoT	IoT (localization)	6G	6G	6G

✗ not included, ✓ included.

[1] Comprehensiveness (Com), Usability (Use), Functionality (Fnc), Robustness (Rob), Neutrality (Neu), Explicability (Exp), Reliability (Rel), Security (Sec), Resilience (Res), Safety (Saf), Privacy (Prv), Availability (Avl), Confidentiality (Cnf), Integrity (Int), Maintainability (Mnt), Adaptability (Adp), Timeliness (Tim), Efficiency (Eff), Reactiveness (Rct), Proactiveness (Pro).

[2] Except [15], the other works do not specify any particular metrics.

[3] The 4-layer protocols in TCP/IP.

[4] Except [15], the other works primarily present generic frameworks and do not specify any particular assessment method.

[5] The binary-level approach provides a binary evaluation for each metric. The coarse-grained approach evaluates using discrete quality levels. The fine-grained approach assigns continuous quality scores for trust assessment.

part of a system's development from the outset, rather than an ad-hoc addition. The authors decompose trustworthiness into six key pillars—security, resilience, privacy, ethics, robustness, and reliability—each evaluated using specific mechanisms and metrics. This structured framework allows the evaluation of trustworthiness with a well-defined rubric while also enabling designers to systematically incorporate and optimize trustworthiness early in the development process.

6G Trustworthiness Assessment Frameworks

In addition to the existing trustworthiness evaluation methods for 6G networks discussed in the literature, several ongoing research efforts aim to develop generic frameworks. 3.1 compares recent frameworks proposed for 6G communications and Internet of Things (IoT) applications. The table indicates that no comprehensive framework currently addresses all technical and non-technical aspects of a trustworthiness assessment workflow for 6G while providing practical implementation guidance and requirements, thereby highlighting future research opportunities.

3.2.3 Challenges and Future Work

Integrating Software-Defined Networking (SDN) in 6G enhances global network trustworthiness by enabling centralized control and visibility, programmability, and automated security orchestration across the core network and Software-Defined Radio Access Networks (SD-RAN). However, 6G introduces new security challenges, including heterogeneous environments, an expanded attack surface, and AI-powered threats, necessitating intelligent and adaptive security measures. Emerging solutions such as secure network slicing, distributed security services under fog-edge-cloud continuum, blockchain-based security, and AI-driven security operations collectively contribute to a more secure and trustworthy 6G network [17, 18]. However, defining and assessing trustworthiness metrics is quite complex in SDN-enabled 6G networks because it requires consideration at all architectural levels. Meanwhile, specialized network monitoring and management software will allow one to accumulate relevant network data and generate appropriate derived metrics.

At the same time, virtual or digital twins are essential for classifying and analyzing trustworthiness metrics in 6G networks [19]. Real-time simulation and modeling of the network with this technology enable testing various threats and their impacts on security, resilience, and trust while refining and verifying trustworthiness metrics. This approach leads to a more accurate metrics assessment and better-informed decision-making in the development and operation of 6G.

In addition, physical layer security (PHY-Sec) is emerging as a key enabler for 6G trustworthiness, leveraging new physical characteristics such as millimeter waves, higher bandwidth, and massive antenna arrays to enhance communication security. The rapid variation in multipath fading and spatial-temporal correlation in 6G channels offers new opportunities for secure key generation and authentication. However, challenges remain in designing robust PHY-Sec techniques that can adapt to dynamic environments and ensure resilience against evolving eavesdropping and spoofing attacks.

Moreover, key enablers such as federated learning, homomorphic encryption, and edge validation for data integrity hold great promise for enhancing privacy in 6G networks. Federated learning assures privacy preservation with decentralized training of machine learning models, while homomorphic encryption enables computations on encrypted data, allowing secure processing of sensitive information and maintaining privacy preservation. Additionally, with the growing number of human-attached sensors in 6G, an automated edge-based validation system is crucial to ensure data integrity of sensors before sharing with applications [1].

However, the mentioned technologies are still in their early research stages and demand further development to meet the requirements of scalability, interoperability, and consistency. Subsequent real-world deployment must consider the complex 6G network trustworthiness challenges. Therefore, future research efforts should focus on developing adaptive trustworthiness architectures and comprehensive assessment frameworks for 6G unified metrics to ensure the seamless integration of these cutting-edge technologies into a trustworthy 6G ecosystem.

3.2.4 References

[1] Ziegler, Volker; Schneider, Peter; Viswanathan, Harish; Montag, Michael; Kanugovi, Sridhar; and Rezaki, Ahmet, "Security and Trust in the 6G Era," *IEEE Access*, vol. 9, pp. 142314–142327, 2021. Available: https://doi.org/10.1109/ACCESS.2021.3120143

[2] Osorio, Diego P. M. Moya; Etzlinger, Bernd; and Karamachoski, Josip, "Trustworthy 6G: Misconceptions, Attributes, and the Labeling Approach," *TechRxiv (preprint)*, 2024. Available: https://d197for5662m48.cloudfront.net/documents/publicationstatus/216343/preprint_pdf/15cfea31b7ee950d5695eb8df0d2e49c.pdf

[3] Fettweis, Gerhard P.; Grünberg, Patricia; Hentschel, Tim; and Köpsell, Stefan, “Conceptualizing Trustworthiness and Trust in Communications,” *arXiv preprint*, 2024. Available: https://doi.org/10.48550/arXiv.2408.01447

[4] Jøsang, Audun; Ismail, Roslan; and Boyd, Colin, “A Survey of Trust and Reputation Systems for Online Service Provision,” *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, 2007. Available: https://doi.org/10.1016/j.dss.2005.05.019

[5] Wang, Yiying; Kang, Xin; Li, Tieyan; Wang, Haiguang; Chu, Cheng-Kang; and Lei, Zhongding, “SIX-Trust for 6G: Towards a Secure and Trustworthy Future Network,” *IEEE Access*, pp.1–1 (early access), 2023. Available: https://doi.org/10.1109/ACCESS.2023.3321114

[6] Ali, Bilal; Gregory, Mark A.; Li, Shancang; and Dib, Omar A., “Zero Trust Security Framework for 5G MEC Applications: Evaluating UE Dynamic Network Behaviour,” in *Proc. 33rd International Telecommunication Networks and Applications Conference (ITNAC)*, Melbourne, Australia, Nov.–Dec. 2023, pp. 140–144. Available: https://doi.org/10.1109/ITNAC59571.2023.10368551

[7] Sharma, Ankur; Pilli, Emmanuel S.; Mazumdar, Amit P.; and Gera, Praveen, “Towards Trustworthy Internet of Things: A Survey on Trust Management Applications and Schemes,” *Computer Communications*, vol. 160, pp. 475–493, 2020. Available: https://doi.org/10.1016/j.comcom.2020.06.030

[8] Sagar, Shabir; Mahmood, Atif; Sheng, Quan Z.; Pabani, Jamil K.; and Zhang, Wei E., “Understanding the Trustworthiness Management in the Social Internet of Things: A Survey,” *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8461–8484, 2022. Available: https://doi.org/10.1016/j.comnet.2024.110611

[9] Peterseil, Philipp; Etzlinger, Bernhard; Horáček, Jan; Khanzadeh, Roya; and Springer, Andreas, “Trustworthiness for an Ultra-Wideband Localization Service,” *Sensors*, vol. 24, no. 16, art. 5268, 2024. Available: https://doi.org/10.3390/s24165268

[10] Trinh-Nguyen, Bao; Berri, Slim; Teo, Siang Guan; Truong-Huu, Tuan; and Chorti, Arsenia, “A Framework for Global Trust and Reputation Management in 6G Networks,” *ResearchGate*, 2024. Available: https://www.researchgate.net/publication/385705828_A_Framework_for_Global_Trust_and_Reputation_Management_in_6G_Networks

[11] Awadid, Afef; Amokrane-Ferka, Kahina; Sohier, Henri; Mattioli, Juliette; Adjed, Faouzi; Gonzalez, Martin; and Khalfaoui, Souhaiel, “AI Systems Trustworthiness Assessment: State of the Art,” in *Proc. Workshop Model-based System Engineering and AI, 12th Int. Conf. Model-Based Software and Systems Engineering (MODELSWARD)*, 2024. Available: https://hal.science/hal-04400795/document

[12] Mo, Jian; Kang, Xin; and Li, Ting, “Toward a Comprehensive Trust Model Assessment for Telecommunication Networks: An Introduction to the Upcoming ITU-T Standard Y.3260,” *IEEE Communications Standards Magazine*, Dec. 2024. Available: https://doi.org/10.1109/MCOMSTD.2024.10802973

[13] Industrial Internet Consortium, “Managing and Assessing Trustworthiness for IIoT in Practice,” Whitepaper, Version 1.0, July 2019. Available: https://www.iiconsortium.org/pdf/Managing_and_Assessing_Trustworthiness_for_IIoT_in_Practice_Whi

tepaper_2019_07_29.pdf#:~:text=Managing%20trustworthiness%20means%20understanding%20the%20trustworthiness%20characteristics%20%28safety%2C,and%20their%20corresponding%20effect%20on%20business%20and%20operations

[14] Liu, Fang; Sun, Ruixuan; Wang, Dong; Javali, Chandrashekhar; and Liu, Peng, "6G Native Trustworthiness," *Communications of Huawei Research*, Sept. 2022. Available: https://www-file.huawei.com/-/media/corp2020/pdf/publications/huawei-research/issue2/6g-native-trustworthiness-en.pdf

[15] Saeedi, Esmaeil S. T.; Valencia, Rafael I. M.; Orozco, Andres L. S.; and Villalba, Luis J. G., "Trust Evaluation Techniques for 6G Networks: A Comprehensive Survey with Fuzzy Algorithm Approach," *Electronics*, vol. 13, no. 3013, 2024. Available: https://doi.org/10.3390/electronics13153013

[16] Bottarelli, Massimiliano; Epiphaniou, Gregory; Mahmood, Shahid; Hooper, Michael; and Maple, Carsten, "Assessing the Trustworthiness of Electronic Identity Management Systems: Framework and Insights from Inception to Deployment," *arXiv preprint arXiv:2502.10771*, 2025. Available: https://arxiv.org/abs/2502.10771

[17] Zuo, Yuhong; Guo, Jing; Gao, Ning; Zhu, Yanhua; Jin, Shi; and Li, Xiang, "A Survey of Blockchain and Artificial Intelligence for 6G Wireless Communications," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2494–2528, 2023. Available: https://doi.org/10.1109/COMST.2023.3315374

[18] Farhoudi, Mohammad; Shokrnezhad, Mohammad; Taleb, Tarik; Li, Ren; and Song, Jing, "Discovery of 6G Services and Resources in Edge-Cloud-Continuum," *IEEE Network*, Early Access, 2024. Available: https://doi.org/10.48550/arXiv.2407.21751

[19] Böck, Yannik N.; Boche, Holger; Schaefer, Rafael F.; Fitzek, Frank H. P.; and Poor, H. Vincent, "Virtual-Twin Technologies in Networking," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 136–141, 2023. Available: https://doi.org/10.1109/MCOM.001.2200861

3.3 Understanding Network Robustness: Measuring Vulnerability and Enhancing Resilience

Authors: Carmela Comito[1], Annalisa Socievole[1]

[1] National Research Council of Italy (CNR), Institute for High Performance Computing and Networking, Italy

3.3.1 Introduction

In an increasingly interconnected world, network resilience has become a critical research area. Network resilience encompasses a system's capacity to withstand, recover from, and adapt to failures, whether they arise from accidental malfunctions, natural disasters, or targeted cyber-physical attacks. The importance of robust network resilience strategies has grown as emerging threats, including sophisticated cyberattacks and large-scale system failures, present unprecedented challenges to infrastructure stability.

In this White Paper, we delve into the critical issue of network robustness. Our focus is on identifying the most crucial links within a network (e.g., wireless telecommunication infrastructures), those whose removal or failure could cause significant disruption. While networks are inherently vulnerable to random failures, they are increasingly at risk from targeted cyber-physical attacks, making it imperative to understand which nodes and links are most vital for maintaining network integrity. The core challenge in enhancing network resilience lies in distinguishing between critical components, whose loss severely impacts functionality, and non-critical elements, which can be removed with minimal effect. Network robustness, at its core, refers to a system's ability to withstand disruptions-whether intentional or accidental-while continuing to operate effectively.

This White Paper provides an overview of network robustness and its role in protecting critical infrastructure systems. We explore the current state-of-the-art, discuss key challenges in network vulnerability assessment, current limitations, and provide future research directions.

3.3.2 State of the Art

Robustness Enhancement via Link Perturbations Improving network robustness often involves topological perturbations, which modify the structure of a network over time (Van Mieghem, 2011). These modifications can take various forms, including the addition or removal of nodes, the introduction or deletion of links, the rewiring of existing connections by changing one of their endpoints, and the adjustment of node and link weights. Targeted perturbations can strengthen or maintain network resilience. Wang et al. (2014) demonstrated both experimentally and theoretically that robustness can be improved by adding links that minimize effective graph resistance, as well as by identifying and protecting links whose removal would significantly increase this resistance. Various strategies have been explored for different network types. In single-layer networks, techniques such as low-degree link addition and random link addition have been proposed to enhance robustness. For interdependent networks, researchers have investigated more advanced strategies, including random inter-degree difference and low inter-degree difference approaches. These methods, along with their extensions, such as low-degree IDD, low-degree-product IDD, and low-degree-sum IDD, have been found to be effective in improving network resilience (Kazawa & Tsugawa, 2020). Several studies have focused on different techniques for perturbation-based robustness enhancement.

Schneider et al. (2011) proposed an iterative rewiring method in which pairs of links are randomly selected and swapped only if the modification improves robustness, ensuring that the overall number of connections remains unchanged. In another approach, Buesser et al. (2011) applied a simulated annealing optimization technique to rewire scale-free networks while preserving the degree distribution, aiming to mitigate the effects of targeted attacks on hub nodes. Carchiolo et al. (2019) adopted a different perspective by enhancing robustness through the addition of a small number of new connections. Unlike most previous approaches, which focus on reinforcing hub nodes, their method establishes long-range backup links between secondary nodes to safeguard network connectivity in the event of hub failures. Furthermore, Louzada et al. (2013) introduced a budget-constrained rewiring strategy based on monitoring the evolution of the network's largest component during targeted attacks. Unlike random rewiring approaches such as those proposed by Schneider et al. (2011), this method guides the formation of a modular onion-like structure, which enhances resilience by organizing nodes into layers based on their degree.

Evolutionary Approaches for Network Robustness In addition to perturbation-based methods, researchers have explored optimization techniques inspired by evolutionary principles to enhance network robustness. Evolutionary computation, a class of optimization techniques that mimic biological evolution (Bäck et al., 1997), has been successfully applied to solve complex real-world optimization problems, including network robustness optimization. These methods initialize a population of potential solutions and iteratively refine them using genetic operators such as mutation, crossover, and selection. By continuously exploring the solution space, evolutionary algorithms can efficiently identify network configurations that improve resilience. Several studies have leveraged evolutionary computation for network robustness optimization. Zhou and Liu (2014) developed a memetic algorithm designed to enhance the robustness of scale-free networks by optimizing the R-value (Herrmann et al., 2011). Their approach focuses on protecting network hubs through targeted rewiring while preserving the overall degree distribution. Similarly, Wang and Liu (2017) proposed a new robustness measure, Rce, which extends the R-value to account for cascading failures. They introduced an evolutionary algorithm, MA-Rce, that integrates genetic operators with a local search mechanism based on simulated annealing, similar to the technique used by Buesser et al. (2011). Building on these approaches, Pizzuti and Socievole (2018, 2019, 2023) developed a series of genetic algorithms for network robustness enhancement. Their first method, RobGA, improves robustness by adding a link that minimizes effective graph resistance. In subsequent work, they introduced RobLPGA, which focuses on protecting network links by identifying and safeguarding those whose removal would most significantly increase

More recently, they addressed the computational challenges associated with robustness optimization by proposing an accelerated version of RobGA. This method leverages an approximation based on the incremental computation of the Moore–Penrose pseudoinverse of the Laplacian matrix, significantly reducing computational complexity while maintaining high accuracy in effective graph resistance calculations.

Enhancing network robustness requires strategic modifications to topology, either through targeted perturbations or optimization techniques. While link perturbations offer immediate, practical improvements, evolutionary methods provide a more adaptive and scalable approach for optimizing network resilience. Future research may further refine these techniques, integrating machine learning and real-time adaptability for enhanced robustness in dynamic networks.

Machine Learning-Based Approaches for Robustness Optimization.

Machine learning (ML) has emerged as a powerful tool for optimizing network robustness by enabling predictive modeling, adaptive decision-making, and intelligent link modifications. Unlike traditional heuristic-based techniques, ML approaches can learn from historical network data, detect patterns of vulnerabilities, and recommend optimal interventions dynamically. This capability is particularly advantageous in large-scale and evolving networks, where robustness requirements change over time.

Graph Neural Networks (GNNs) for Robustness Prediction GNNs have gained popularity as an effective ML framework for analyzing network structures [1]. Unlike traditional neural networks, GNNs can directly process graph data, making them well-suited for predicting network vulnerabilities and identifying optimal link modifications to enhance robustness. By encoding the relationships between nodes and learning hierarchical features from the network topology, GNNs can predict which structural modifications will most effectively improve resilience. Several studies have demonstrated the potential of GNNs in network robustness optimization. In [1] GNN models have been used to analyze network vulnerability patterns and predict the best reinforcement strategies against targeted attacks. Their work showed that deep learning models could outperform traditional heuristics by adapting to complex network structures and providing data-driven recommendations for improving robustness.

Reinforcement Learning (RL) for Adaptive Robustness Optimization RL provides another promising avenue for robustness enhancement by modeling network optimization as a sequential decision-making process. Unlike static optimization techniques, RL allows an intelligent agent to interact with the network, learn from its responses, and iteratively refine its strategy for maximizing robustness. This approach is particularly beneficial for dynamic networks, where the topology evolves over time due to external factors such as traffic fluctuations, environmental changes, or cyberattacks. Recent studies have explored the application of RL in network resilience. In [2] is proposed an RL-based framework for enhancing the robustness of interdependent networks by optimizing link allocation. Their approach allowed the network to adaptively reinforce its structure in response to observed failures, improving resilience against cascading disruptions.

Anomaly Detection for Network Resilience Another important ML application in network robustness is anomaly detection, which involves identifying unusual patterns that may indicate structural weaknesses or impending failures. By analyzing historical network performance data, ML classifiers can detect deviations from normal behavior and trigger proactive interventions to reinforce network resilience.

Link Perturbation Methods for Robustness Enhancement

Enhancing the robustness of networks is often achieved by introducing perturbations to their topology. A perturbation represents an event that alters the network structure over time, consisting of a sequence of fundamental modifications. These fundamental changes, occurring at specific time intervals, influence the underlying graph representation of the network, affecting matrices such as the adjacency matrix or the Laplacian matrix [5]. The modifications can take various forms, including the addition or removal of nodes, the introduction or elimination of links, the reassignment of link endpoints (rewiring), and changes in either node or link weights.

By carefully designing these perturbations, network robustness can be actively preserved or even improved. For instance, when considering link perturbations in relation to the concept of effective graph resistance as a robustness indicator, Wang et al. [5] provide both theoretical and empirical evidence that two key strategies enhance network resilience: (i) strategically adding links that minimize effective graph resistance and (ii) identifying and safeguarding critical links whose removal would significantly increase this resistance. Their study assesses four different link selection strategies across a variety of real and synthetic networks, quantifying the impact of these structural changes on overall robustness. In cases where networks are subjected to degree-based targeted attacks, the addition of specific links has been demonstrated to be a viable defense strategy, particularly in interdependent and multilayered networks. Kazawa and Tsugawa [6] analyze the effectiveness of multiple link-addition strategies, distinguishing between methods for single-layer networks (such as low-degree (LD) and random addition (RA)) and those tailored for interdependent networks.

Beyond link additions, network rewiring has also emerged as a powerful tool for increasing robustness. Schneider et al. [7] propose an iterative rewiring process where random link pairs are selected, and swaps are performed only if they result in an improvement in network resilience. This method ensures enhanced robustness without altering the total number of connections. Similarly, in the context of scale-free network design, a structural configuration resembling an "onion-like" topology has been suggested, in which high-degree nodes are centrally located while nodes with progressively lower degrees form the outer layers, enhancing overall resilience. Instead of reinforcing highly connected nodes, some studies explore alternative strategies for enhancing network robustness. Carchiolo et al. [8] propose introducing a small number of additional links to scale-free networks, deliberately connecting nodes that play a secondary role rather than targeting hubs. This approach aims to establish long-range alternative paths, acting as backup connections in the event of failures in central nodes. Unlike traditional strategies that reinforce the most influential components of a network, this method fosters resilience by ensuring the existence of redundant pathways even in the periphery of the system.

Percolation theory provides a theoretical framework for studying how networks maintain connectivity under progressive failures [4]. By analyzing the behavior of networks as nodes or links are removed, percolation-based methods offer valuable insights into structural resilience and provide strategies for reinforcing weak components. Bootstrap percolation is a probabilistic model in which nodes fail if they do not have a sufficient number of active neighbors. This cascading process helps identify regions of the network that are particularly vulnerable to fragmentation, allowing targeted reinforcement strategies to be applied. Strengthening these weakly connected regions can significantly enhance overall robustness.

Cascading Failure Mitigation Strategies

Cascading failures represent one of the most severe threats to network robustness, particularly in interdependent and multilayer networks [9]. A small initial failure can propagate throughout the system, leading to large-scale disruptions. Developing strategies to mitigate cascading failures is essential for maintaining the stability of critical infrastructure networks. One effective approach for mitigating cascading failures is to dynamically redistribute load across the network to prevent overload conditions.

Interdependent networks, such as power grids and communication systems, require strategies to dynamically adjust their coupling strength to prevent systemic collapse. When dependencies between networks are too rigid, failures in one system can rapidly propagate to the other. Adaptive coupling techniques aim to regulate these dependencies to enhance resilience.

Another important strategy for mitigating cascading failures is the containment of failing network regions. By isolating affected nodes or subnetworks, further disruptions can be prevented.

3.3.3 Challenges and Future Work

While significant progress has been made in enhancing network robustness, several challenges remain. These challenges stem from computational complexity, real-world applicability, adaptability to dynamic environments, and trade-offs between robustness and other network properties. Addressing these issues will pave the way for more effective and scalable robustness enhancement strategies.

Many robustness optimization problems, including identifying the most critical links to add or protect, are computationally intractable due to their combinatorial nature. New optimization techniques leveraging graph sparsification, distributed computing, or quantum computing could help improve computational efficiency in large-scale networks.

Many networks are not isolated but interdependent, meaning failures in one network can trigger cascading failures in another. Future studies should investigate how robustness measures and optimization techniques can be extended to multilayer and interconnected networks, where dependencies between layers introduce additional vulnerabilities. Future research in this area could explore hybrid approaches that combine bootstrap percolation with ML-based predictive models to develop more adaptive robustness enhancement strategies.

By addressing these challenges and research gaps, future work can develop more efficient, adaptive, and practical robustness optimization strategies, making networks more resilient against failures and adversarial disruptions.

3.3.4 References

[1] Mitra, Shaswata; Chakraborty, Trisha; Neupane, Subash; Piplai, Aritran; and Mittal, Sudip, "Use of Graph Neural Networks in Aiding Defensive Cyber Operations," *arXiv preprint*, arXiv [cs.CR], 2024. Available: https://doi.org/10.48550/arXiv.2401.05680

[2] Lotfi, Fatemeh; and Afghah, Fatemeh, "Meta Reinforcement Learning Approach for Adaptive Resource Optimization in O-RAN," in *Proc. 2025 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, 2025. Available: https://doi.org/10.1109/WCNC61545.2025.10978365

[3] Kumar, D.; Routhu, S.; Manjula, B.; and Routhu, S., "Anomaly Detection in Networks Using Machine Learning Techniques," *International Journal of Natural Sciences*, 2025. Available: https://www.researchgate.net/profile/Routhu-Daswanta-Kumar/publication/388174855_Anomaly_Detection_in_Networks_using_Machine_Learning_Techniques/links/678ea9a695e02f182ea58ae7/Anomaly-Detection-in-Networks-using-Machine-Learning-Techniques.pdf

[4] Louzada, Vinicius H. P.; Daolio, Fabio; Herrmann, Hans J.; and Tomassini, Marco, "Smart Rewiring for Network Robustness," *Journal of Complex Networks*, vol. 1, no. 2, pp. 150–159, 2013. Available: http://dx.doi.org/10.1093/comnet/cnt010

[5] Wang, Xiangrong; Pournaras, Evangelos; Kooij, Robert E.; and Van Mieghem, Piet, "Improving Robustness of Complex Networks via the Effective Graph Resistance," *The European Physical Journal B*, vol. 87, no. 9, article 221, Sept. 2014. Available: https://doi.org/10.1140/epjb/e2014-50276-0

[6] Kazawa, Yui; and Tsugawa, Sho, "Effectiveness of Link-Addition Strategies for Improving the Robustness of Both Multiplex and Interdependent Networks," *Physica A: Statistical Mechanics and its Applications*, vol. 545, article 123586, May 2020. Available: https://doi.org/10.1016/j.physa.2019.123586

[7] Schneider, Christian M.; Moreira, André A.; Andrade Jr., José S.; Havlin, Shlomo; and Herrmann, Hans J., "Mitigation of Malicious Attacks on Networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 10, pp. 3838–3841, Mar. 2011. Available: https://doi.org/10.1073/pnas.1009440108

[8] Carchiolo, Vincenza; Grassia, Marco; Longheu, Alessandro; Malgeri, Michele; and Mangioni, Giuseppe, "Network Robustness Improvement via Long-Range Links," *Computational Social Networks*, vol. 6, no. 1, article 12, Dec. 2019. Available: https://doi.org/10.1186/s40649-019-0073-2

[9] Xing, Liudong, "Cascading Failures in Internet of Things: Review and Perspectives on Reliability and Resilience," *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 44–64, Jan. 2021. Available: https://doi.org/10.1109/JIOT.2020.3018687

3.4 Trustworthiness-Aware Network Management in 6G Communications

Authors: Roya Khanzadeh[1,2], Fjolla Ademaj-Berisha[3,2] and Andreas Springer[1,2]

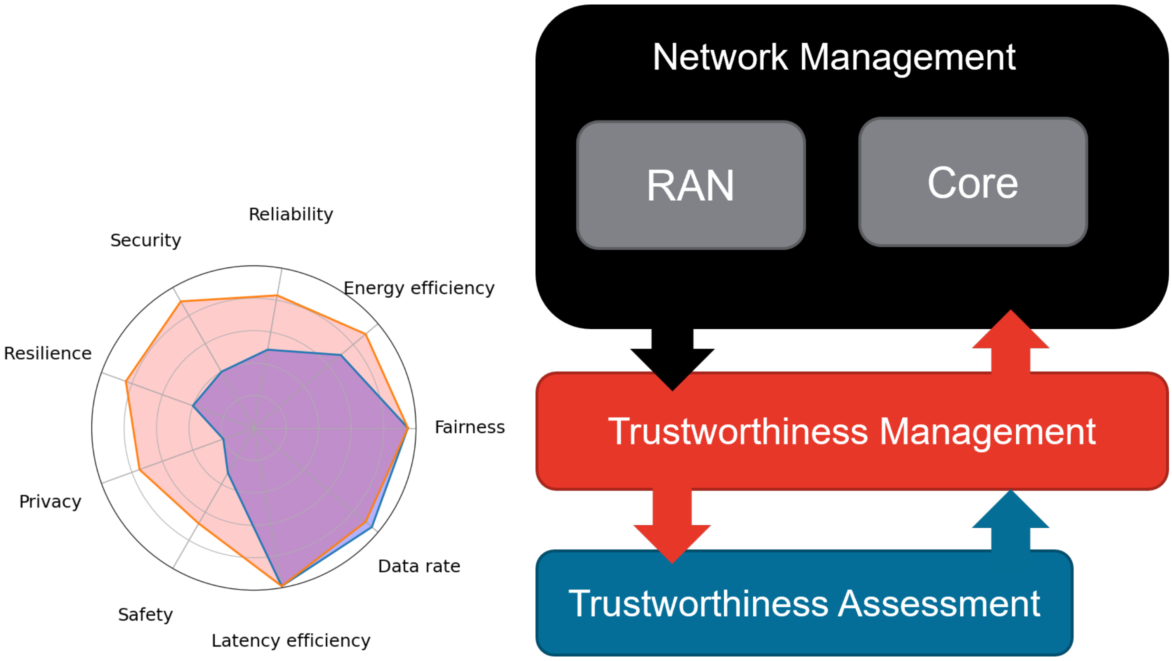
[1]Johannes Kepler University Linz, Institute for Communications Engineering and RF-Systems, Linz, Austria
[2]JKU LIT SAL IWS Lab, Linz, Austria
[3]Silicon Austria Labs GmbH, Linz, Austria

3.4.1 Introduction

The sixth generation (6G) of mobile networks is envisioned to deliver groundbreaking services while generating massive amounts of data and utilizing vast network resources, increasingly depending on trustworthy network operations. A trustworthy 6G network must consistently perform as expected, incorporating key characteristics such as reliability, security, resilience, privacy, and safety. Reliability in 6G refers to the network's ability to consistently deliver data packets with minimal errors, latency, or faults under diverse conditions. Security ensures that communication links remain protected from eavesdropping and tampering, safeguarding data integrity and confidentiality. Resilience focuses on maintaining network availability despite potential threats, requiring robust mitigation and recovery mechanisms. Privacy emphasizes that users must retain control over their personal and location data, ensuring not only anonymity and confidentiality but also integrity, availability, and compliance with the principle of minimality. Finally, safety is the ability of the network to ensure that mobile network operations do not lead to direct/indirect catastrophic impacts on human life, health, property, data, or the surrounding physical environment [1].

A trustworthy 6G network requires both the evaluation/assessment and enhancement of trustworthiness, considering the aforementioned characteristics. This necessitates continuous assessment and active management throughout network operation. While trustworthiness requirements are factored into the design phase of mobile networks, the overall trustworthiness of a 6G system largely depends on operator-driven deployment optimization, implementation, and configuration. In the 6G era, networks should leverage zero-trust principles [2] and incorporate a dynamic trust model, enabling real-time trustworthiness assessment and adaptation. This requires continuous verifying and tracking users' legitimacy and behavior, and monitoring network links, as well as ongoing analysis of network subsystems' behavior, conditions, and requirements. Based on this evaluation, the trust level of users or communication links can be dynamically adjusted, upgraded or downgraded, when accessing network resources. Additionally, as 6G networks become more flexible and scalable, trustworthiness mechanisms must be fine-grained and dynamically adaptable to evolving network conditions, user demands, and business requirements.

Recently, the question has arisen as to whether a dedicated trustworthiness management layer should be incorporated into 6G network operations. Such a layer would be responsible for ensuring and enhancing trustworthiness across key network management components, including the radio access network (RAN) and Core network, by optimizing tasks such as scheduling, resource allocation and slicing, and routing [3]. A trustworthiness-aware network management layer must strike a balance between trustworthiness metrics and traditional network key performance indicators (KPIs). Achieving this requires an adaptive network architecture capable of dynamic decision-making in network management tasks. For instance, it must determine when and where to apply security measures, such as encryption or physical-



(a) (b)

Figure 3.2: (a) Conceptual model for trustworthiness-aware network management. (b) Radar chart comparing traditional KPI-based network management performance (blue line) vs. that of the trustworthiness-enhanced network management (red line).

layer security (PHYSec), and how to distribute resources (e.g., power and bandwidth) to maintain trustworthiness in terms of security, reliability, and privacy, while simultaneously optimizing KPIs such as data rate, throughput, and delay. Figure 3.2 illustrates a conceptual model for integrating trustworthiness into network management, demonstrating how it potentially can enhance trustworthiness while balancing traditional network KPIs. Although the integration of trustworthiness management into network operations is still an evolving concept in 6G networks, several existing methods in the literature have explored enhancing specific trustworthiness characteristics through network management tasks, such as resource allocation, routing, task offloading, and scheduling, as discussed in the following section.

3.4.2 State of the Art

Security considerations play a critical role in resource allocation decisions, as improper allocation can compromise the network's ability to prevent, detect, and mitigate security threats. Security has been even accounted as one of the KPIs for network-aware resource management in 6G [4]. Security can be integrated into resource allocation either as an optimization objective or as a constraint to ensure a network's resilience against attacks and failures. Insufficient computational power may prevent the execution of necessary encryption and decryption tasks, exposing the network to eavesdropping and data breaches. Similarly, inadequate bandwidth may limit the network's capacity to detect and mitigate distributed denial of service (DDoS) or side-channel attacks, leading to service disruptions. Moreover, in distributed systems, resource allocation strategies must ensure there is no single point of failure, enabling the network to maintain service continuity through redundancy and failover mechanisms [5]. In [6], the problem of security-aware resource allocation is addressed by proposing a side channel-aware resource allocation algorithm for ultra-reliable low-latency communication (URLLC) and enhanced mobile broadband (eMBB) slices in 5G RAN. With the constraint of avoiding side-channel attacks, the objective is to maximize the number of slices accommodated in 5G RAN by optimizing resource allocation for slices.

On the other hand, in wireless communication scenarios where direct communication is not possible for all nodes, network trustworthiness might be compromised, as making reliable data transmission and timely response to critical events is challenging. Addressing this issue requires innovative strategies that optimize routing paths while considering trustworthiness characteristics such as reliability and security. A learning-based secure routing algorithm is proposed in [7] to enhance security in edge networks under dynamic DoS attacks. The approach predicts attack patterns by analyzing both historical and real-time data, enabling the selection of secure routing paths that minimize the risk of interception and maximize the probability of successful data transmission.

Additionally, task offloading is one of the network management tasks in 6G vehicular services, where vehicles rely on edge access points (e.g., roadside units) for computational tasks. However, the zero-trust paradigm requires dynamic and secure task allocation. To mitigate risks from malicious edge nodes, reputation-based offloading strategies have been explored in [8] to assess the reputation of edge nodes, ensuring tasks are offloaded only to trusted nodes. Additionally, a federated asynchronous reinforcement learning algorithm is employed to optimize offloading decisions, enhancing both security and network performance.

Moreover, scheduling in wireless networks is traditionally designed to optimize performance metrics such as bandwidth, power, latency, throughput, and fairness. However, in 6G networks, scheduling must also balance trustworthiness metrics alongside these traditional criteria. A reinforcement learning-based scheduler is introduced in [9], that leverages environmental knowledge, the contextual awareness of network conditions, including channel state and mobility pattern, to enhance reliability and security. This approach defines service reliability as a measure of network consistency and service availability as an indicator of security, as the lack of service availability has consequences for security breaches. The proposed reinforcement learning-based scheduler finds a balance between network trustworthiness (in terms of reliability and security) and fairness among the users.

While the above-mentioned works address trustworthiness in network management, they mostly focus on limited aspects of trustworthiness, highlighting the need for a more holistic integration of trustworthiness into network management task optimization.

3.4.3 Challenges and Future Work

Many of the technological advancements in 6G will build upon existing foundations from previous generations. However, trustworthiness has recently emerged as a key concern in 6G communication, driven by its diverse applications and increasing security, reliability, and privacy demands. Therefore, several critical challenges remain largely unexplored in the integration of trustworthiness in 6G network management and must be addressed to ensure the dependable operation of 6G networks. Some of the key challenges and future research directions in this domain are outlined below:

- How can trustworthiness assessment and management be seamlessly integrated into 6G networks to enable continuous evaluation and adaptive enhancement of reliability, security, resilience, privacy and safety?
- What are the most effective methods for trustworthiness-enhanced network management tasks such as scheduling, resource allocation, and routing? How to improve trustworthiness as much as possible without compromising the network KPIs?

- How can environmental awareness and sensing data be leveraged in 6G networks to enhance trustworthiness by mitigating not only external attacks but also errors, failures, and vulnerabilities inherent in stochastic wireless channels?
- How can 6G networks balance security and privacy in positioning, localization, and tracking, while leveraging sensing data to enhance trustworthiness without enabling unauthorized tracking or malicious control of human and non-human assets?
- How to coordinate user-centric safety in network optimization while addressing specific trustworthiness requirements and data governance policies specified by users throughout the 6G networks?

Acknowledgement

This work has been in part supported by the "University SAL Labs" initiative of Silicon Austria Labs (SAL) and its Austrian partner universities for applied fundamental research for electronic-based systems.

3.4.4 References

[1] NGMN Alliance, *6G Trustworthiness Considerations*, Next Generation Mobile Networks Alliance, 2023. Available: https://www.ngmn.org/publications/6g-trustworthiness-considerations.html

[2] Chen, Xu; Feng, Wei; Ge, Ning; and Zhang, Yan, “Zero Trust Architecture for 6G Security,” *IEEE Network*, 2023. Available: https://doi.org/10.1109/MNET.2023.3326356

[3] Fettweis, Gerhard P.; and Boche, Holger, “On 6G and Trustworthiness,” *Communications of the ACM*, vol. 65, no. 4, pp. 54–61, 2022. Available: https://doi.org/10.1145/3512996

[4] Sefati, Seyed Salar; Haq, Asim Ul; Craciunescu, Razvan; Halunga, Simona; Mihovska, Albena; Fratu, Octavian; *et al.*, “A Comprehensive Survey on Resource Management in 6G Network Based on Internet of Things,” *IEEE Access*, 2024. Available: https://doi.org/10.1109/ACCESS.2024.3444313

[5] Khan, Md Muhidul Islam; Islam, Md Tareq; Al-Fuqaha, Ala; *et al.*, “Resource Allocation in Networking and Computing Systems: A Security and Dependability Perspective,” *IEEE Access*, 2023. Available: https://doi.org/10.1109/ACCESS.2023.3306534

[6] Li, Yajie; Zhao, Yongli; Li, Jun; Zhang, Jiawei; Yu, Xiaosong; and Zhang, Jie, “Side-Channel Attack-Aware Resource Allocation for URLLC and eMBB Slices in 5G RAN,” *IEEE Access*, vol. 7, pp. 2090–2100, 2019. Available: https://doi.org/10.1109/ACCESS.2019.2962179

[7] Wang, Xiaolin; Chen, Cailian; He, Jianping; Zhu, Shanying; and Guan, Xinping, “Learning-Based Online Transmission Path Selection for Secure Estimation in Edge Computing Systems,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 7293–7303, 2020. Available: https://doi.org/10.1109/TII.2020.3012090

[8] Hao, Min; Ye, Dongdong; Wang, Siming; Tan, Beihai; and Yu, Rong, “URLLC Resource Slicing and Scheduling for Trustworthy 6G Vehicular Services: A Federated

Reinforcement Learning Approach," *Physical Communication*, vol. 48, 2021. Available: https://doi.org/10.1016/j.phycom.2021.101470

[9] Ademaj-Berisha, Fjolla; Khanzadeh, Roya; Springer, Andreas; and Bernhard, Hans-Peter, "Environmental-Aware Reinforcement Learning-Based Scheduler for Trustworthy 6G in the Factory Floor," *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2024. Available: https://doi.org/10.1145/3636534.3696728

3.5 The (missing) Privacy Links Between NIS2, CRA and GDPR and their Impact on 6G Networks

Authors: Ana Ferreira[1], Başak Ozan Özparlak[2]

[1] RISE-Health, MEDCIDS, Faculty of Medicine, University of Porto, Portugal
[2] Faculty of Law, Ozyegin University, Istanbul, Türkiye

3.5.1 Introduction

The European Union (EU) has been introducing several regulations and directives with the aim of achieving and maintaining a high common level of privacy and security within the Union. However, a clear and harmonized view of those obligations is harder to envisage, and this can be prone to missing links and blind spots that need to be identified as soon as possible. This section of the White Paper aims to explore and discuss how NIS2 [2] and CRA [2] must integrate privacy by design requirements from GDPR [3] and highlight the main (missing) links between these regulations and their impact on 6G networks [4]. This work contributes to the harmonization of several legal and privacy requirements, not only to increase literacy, but also to promote a call for action for faster and more adequate compliance with the various EU regulations, in practice.

3.5.2 State of the Art

NIS2 is a Directive aiming to bring technical and methodological obligations for ex-ante cybersecurity risk management measures of entities that provide critical infrastructure services in the EU. NIS2 was adopted in 2022 and is effective in all member states from 18th October 2024. NIS2 is the natural evolution from NIS, to provide alignment with the technological breakthroughs and increased interconnectivity that will be introduced by emerging technologies or the 6G network and beyond. There are ten measures defined in NIS2, Art. 21 (2), which comprises the minimum obligations to integrate legality by design into the development and management of entities' network and information systems infrastructure. But do these measures comprise requirements from privacy by design (PbD) as mandated by GDPR? PbD means that privacy safeguards must be included by default into networked data systems and technologies [5]. In the sense of 6G, according to PbD, privacy must be incorporated into the 6G network, including all the enabling technologies, by default. NIS Recital (51) has a direct reference as to how GDPR requirements for data protection by design and by default must be "fully exploited". While NIS2 promotes full cybersecurity processes and policies to support and backup those processes regarding risk and cybersecurity management, GDPR PbD principles mostly integrate scattered and general measures that can be picked at each entity's will, such as minimization, purpose limitation, pseudonymization or accountability, to name a few. These measures are not clearly integrated into more complex processes, such as Incident handling or Business Continuity and Disaster Recovery Plans, as mandated by NIS2. Also relevant is a link missing between NIS2 Article 21(2.g) "basic cyber hygiene practices and cybersecurity training" and GDPR PbD principles. Training and awareness can be successful vehicles for addressing human factor vulnerabilities while promoting prevention of personal data breaches. However, no clear advice on this issue is referred in GDPR.

Another cybersecurity legislation in EU, the Cyber Resilience Act (CRA), entered into force on 10 December 2024 to provide a common cybersecurity framework from the design phase of products with digital elements, needs also to be explored in terms of PbD needs, as

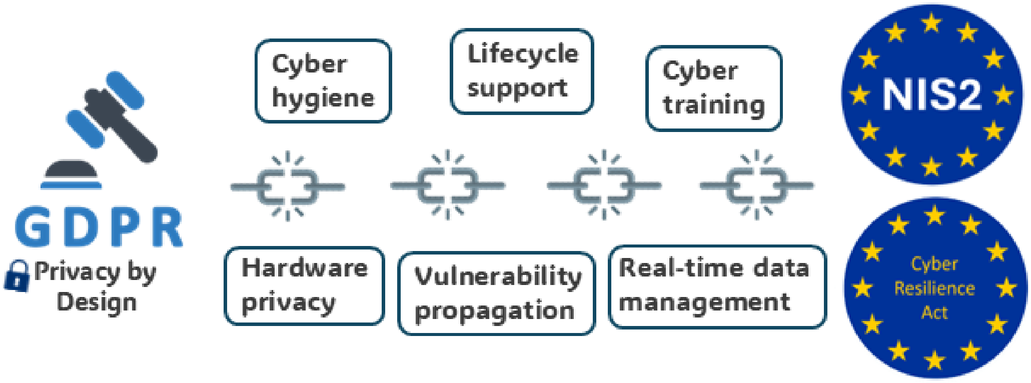


Figure 3.3: The missing privacy links between NIS2, CRA and GDPR.

it aims to enforce products to be secure by design (SbD). CRA aims to allow users to choose products with adequate cybersecurity properties or use them in a secure manner (Recital (1)). Hence, CRA addresses information asymmetry in favor of the consumers by enforcing mandatory security standards. CRA is applicable both for hardware and software products and will be fully enforced in 2027. According to CRA Recital (9), malicious actors can attack all digital items incorporated into or connected to a broader electronic information system under specific conditions. Thus, even less critical hardware and software can help compromise a device or network, allowing hostile actors to obtain privileged access or migrate laterally. Manufacturers must design and build all connectable goods with digital features to meet the CRA's key standards of SbD. As such, CRA is also closely interconnected with GDPR requirements of PbD. It is not possible to have a cybersecurity resilient system without it also integrating privacy throughout its lifecycle. CRA Recital (32) expresses the contribution to GDPR enforcement with the establishment of data protection certification mechanisms and of data protection seals and marks, for the purpose of demonstrating compliance of processing operations by controllers and processors with GDPR. Moreover, NIS2 and CRA promote cybersecurity iteratively over products and services' lifecycle. It is not clear how, within the supply chain, GDPR can give directions to tackle complex vulnerability management such as the cascading or propagation of privacy vulnerabilities as stated in Recitals (9) and (43) for the various system's hardware and software components (Figure 3.3).

Neither NIS2 nor CRA explicitly detail how PbD should be implemented at a hardware level or how to tackle real-time data management in a hyperconnected network like 6G. Apparently, those regulations (and even the EU AI Act [6]) trust and delegate their privacy needs to GDPR. Next, a discussion on the challenging links between NIS2, CRA and GDPR in relation to 6G is provided.

3.5.3 Challenges and Future Work

Since NIS2, CRA, and GDPR will constitute a fundamental basis for cybersecurity in 6G networks, efforts to ensure alignment between these three legislations are also crucial for the effectiveness of the AI Act and other relevant regulations for 6G. It is also timely to tackle the legislative gaps for two reasons. First, even if it is unclear for now, a chance for reform in GDPR is in the air; and secondly, the rapid developments in Generative AI (GenAI) and the impacts on both privacy and security of embedding Large Language Models (LLMs) into 6G network has just started to be discussed. In a technical turbulence like this, it is also inadequate for a data protection regulation focusing on the protection of personal data, like

GDPR, to not provide sufficient protection for privacy or security in the 6G network, as applications that will be needed by the new communication techniques that will come with 6G, will need the data not only from the users but also from all the elements in the users' environment. This will lead to the inadequacy of the current data protection law focused on the protection of personal data because it will potentially be more difficult to distinguish between personal and non-personal data. Another situation that can reduce the effectiveness of resilience is that the speed of the benefits of legal remedies could be slower compared to the speed of data processing in 6G and its consequences. This implies the importance of proactive and adequate approaches such as PbD for ensuring privacy and security in next-generation networks.

Given the significant role that AI will play in network operations in the 6G era and since NIS2 promotes the use of AI to strengthen cybersecurity (Recital 51), understanding the connection between GDPR, NIS2, and CRA along with identifying any gaps before the 6G network is commercialized is crucial. The dependency of the 6G network on AI can bring better security and privacy measures [7], but it can also increase risks against those same measures. The vulnerability surface will rise when large language models (LLMs) are introduced into the 6G network [10]. As highlighted in the International AI Safety Report, the key issue in privacy protection for GenAI arises from privacy-noncompliance which, most of the time, cannot be identified until a data breach occurs [8]. We hereby identify this as hidden by design. This makes PbD significantly more essential, especially for AI-integrated 6G networks because, as a proactive method, PbD anticipates and prevents privacy risks, which are often tied to security breaches, before they occur. Being powered by the capacities of generative AI, 6G network will be data-centric in a way that has never been seen before. Without robust and harmonized security measures, privacy cannot be effectively ensured, particularly in the context of next-generation wireless networks like 6G, as "without strong security, there can be no privacy". This is the reason that more work on privacy and security specific to 6G networks is required for the harmonized and effective application of NIS2, CRA and GDPR. To enable 6G networks to be secure, reliable and trustworthy, the 6G Industry Association (6G-IA) underlines the importance of merging emerging technologies with user-centric principles [9] like PbD. However, understanding the correlation and gaps between EU legal frameworks on security and privacy is a must for harmonized 6G network security. Software dependencies for all industries, including supply chain security, are among the top emergency concerns for 2030 by ENISA [11]. NIS2 and CRA compliance should be interpreted via the framework of GDPR, particularly with regards to gaps for privacy by design principles to enhance resilience. This would leverage HEXA-X-II's approach "to embed security, privacy, and resilience measures throughout the network architecture" [12].

3.5.4 References

[1] European Union, "Directive (EU) 2022/2555 – NIS2 Directive," 2022. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022L2555

[2] European Union, "Regulation (EU) 2024/2847 – Cyber Resilience Act (CRA)," 2024. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R2847

[3] European Union, "Regulation (EU) 2016/679 – General Data Protection Regulation," 2016. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[4] European Commission, "6G," Shaping Europe's Digital Future, 2024. Available: https://digital-strategy.ec.europa.eu/en/policies/6g

[5] Cavoukian, Ann, "Privacy by Design: The 7 Foundational Principles," White Paper, Information and Privacy Commissioner of Ontario, Canada. Available: https://privacy.ucsc.edu/resources/privacy-by-design---foundational-principles.pdf

[6] European Union, "Regulation (EU) 2024/1689 – Artificial Intelligence Act (AI Act)," 2024. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

[7] Porambage, Pradeep; Gür, Gürkan; Moya Osorio, Diego P. M.; Liyanage, Madhusanka; and Ylianttila, Mika, "6G Security Challenges and Potential Solutions," in *Proc. Joint European Conference on Networks and Communications & 6G Summit*, Porto, Portugal, 2021, pp. 622–627. Available: https://doi.org/10.1109/EuCNC/6GSummit51104.2021.9482609

[8] Department for Science, Innovation and Technology UK, "International AI Safety Report," 2025. Available: https://www.gov.uk/government/publications/international-ai-safety-report-2025

[9] 6G-IA Security Working Group, "Innovative Approaches for 6G Security," Position Paper, 2025. Available: https://6g-ia.eu/wp-content/uploads/2025/01/wg_sec_position_paper-23.pdf

[10] Loven, L.; Bordallo Lopez, M.; Morabito, R.; Sauvola, J.; and Tarkoma, S. (Eds.), *Large Language Models in the 6G-Enabled Computing Continuum: A Whitepaper*, 6G Research Visions No. 14, University of Oulu, 2025. Available: https://urn.fi/URN:NBN:fi:oulu-202501211268

[11] ENISA, "2024 Report on The State of Cybersecurity in the Union," Whitepaper, 2024. Available: https://enisa.europa.eu/publications/2024-report-on-the-state-of-the-cybersecurity-in-the-union

[12] Wendt, S. (Ed.), *Environmental, Social, and Economic Drivers and Goals for 6G*, Hexa-X-II Deliverable D1.1, Version 1.2, Hexa-X-II Consortium, June 2023. Available: https://hexa-x-ii.eu/wp-content/uploads/2023/07/Hexa-X-II_D1.1_final-website.pdf

3.6 Supply Chain Cybersecurity in 5G/6G networks: Risks, Regulations, and Beyond

Authors: Gergely Biczók[1,2], András Gazdag[1]

[1] CrySyS Lab, Budapest University of Technology and Economics, Hungary
[2] HUN-REN-BME Information Systems Research Group, Hungary

3.6.1 Introduction

Next Generation Mobile Networks will be composed of integrated software, hardware, and cloud services from multiple specialized vendors rather than a single supplier; in fact, some 5G networks are already being deployed along these lines. The interdependencies within this supply chain pose significant cybersecurity risks, as mobile network operators depend on the security practices of numerous entities beyond their direct control [1]. Recent cyberattacks have exploited vulnerabilities within these supplier networks, highlighting the supply chain as a critical weakness [5].

On a related note, the European Union has already recognized the importance of supply chains in digital products and services and critical infrastructure. The recent NIS2 [2] and Cyber Resilience Act (CRA) [3] define several requirements to properly manage supply chain risks both on the organization and product/service level. Furthermore, other technology-related regulations could also affect the compliance requirements of equipment manufacturers, software vendors, network operators, and other stakeholders.

This paper aims to i) map out supply chain risks in current 5G and foreseen 6G networks ii) investigate these through the lens of European regulation, iii) assess whether a risk-based legal framework is satisfactory for improving supply chain cybersecurity, and finally iv) provide a glimpse of the situation outside Europe.

3.6.2 State of the Art

Cybersecurity risks in the 5G/6G supply chain

The security of 5G networks is deeply intertwined with the complexity of their supply chains, where vulnerabilities can emerge at multiple levels. The hardware supply chain consists of core manufacturers, software developers, security providers, and cloud services at Tier 1, extending through subcontractors and suppliers across multiple tiers. At each stage, risks such as counterfeit components, malicious code, backdoors, and insider threats can compromise the integrity of the network. These vulnerabilities introduce potential points of failure that adversaries can exploit, whether through tampered hardware, unauthorized access, or embedded security flaws [1].

Compounding these hardware risks, the software development pipeline in 5G networks presents additional attack vectors. Threat actors target collaborative repositories, build engines, and code repositories, injecting malicious code through compromised third-party tools, infected testing environments, and manipulated updates. This results in security breaches that can propagate through the continuous integration (CI) and continuous deployment (CD) processes, ultimately affecting deployed 5G systems. These risks are particularly concerning as attackers can exploit them to introduce undetected vulnerabilities, steal sensitive data, or disrupt network operations.

With both hardware and software vulnerabilities deeply embedded in the 5G ecosystem, identifying and assessing threats becomes increasingly difficult. The layered nature of the supply chain obscures accountability, making it challenging to track the origin of security breaches. Moreover, as attackers leverage sophisticated methods-ranging from exploiting supply chain dependencies to injecting malware into critical infrastructure-the overall security posture of 5G networks remains under constant threat. Without visibility into these interconnected risks, vulnerabilities persist, leaving networks exposed to potential compromise at every stage of development and deployment.

The road to 6G. The evolution of networks towards 6G will bring new attack surfaces [9]. First, the predicted 6G architecture will integrate Non-Terrestrial Network (NTN) segments such as Low Earth Orbit (LEO) satellite constellations, High Altitude Platform Stations (HAPS), and Uncrewed Aerial Vehicles (UAVs) to boost global connectivity, to serve rural areas and airline passengers, and to provide situational flexibility for emergency preparedness. Second, 6G is foreseen to be ML-native, relying on the architecture's inherent ML capabilities for dynamic resource allocation, increased resiliency (including security), and energy efficiency. However, the increasing reliance on ML models can also magnify the novel risks concerning ML models, such as adversarial examples, model poisoning, prompt injections, sponge attacks, and so on [11]. Both of these factors increase the importance of supply chain security as new NTN network nodes and ML models are unlikely to be developed and manufactured by the network operators (or their traditional equipment vendors) themselves.

European regulation

As supply chain cybersecurity questions have started to garner attention, the European Union has made it a focal point in its new cybersecurity-related regulations. Chief among these for network equipment and software vendors are the Cyber Resilience Act and the new Product Liability Directive.

Cyber Resilience Act, 2024. The Cyber Resilience Act (CRA) [3], which took effect on December 10, 2024, establishes mandatory cybersecurity requirements for manufacturers and retailers of digital products, referred to as Product Digital Elements (PDEs). The regulation enforces built-in security measures throughout the entire product lifecycle (a principle known as the *duty of care*) to address cybersecurity vulnerabilities and provide consumers with a standardized framework for identifying secure products.

The CRA classifies PDEs into three risk-based categories. Low-risk products, such as smart speakers, must adhere to fundamental compliance requirements through self-assessment. "Critical Class I" products, including password managers, are subject to more stringent security standards, which necessitate conformity with recognized ISO/IEC/ETSI certifications or independent audits. The highest-risk PDEs, by contrast, require mandatory third-party audits. Compliance with the CRA is obligatory; however, PDEs that are already certified under the voluntary European Cybersecurity Certification Scheme (EUCC) automatically fulfill CRA requirements.

A key focus of the CRA is supply chain security, particularly concerning third-party components incorporated into PDEs. Component manufacturers selling within the EU must comply with the regulation, while PDE manufacturers bear responsibility for ensuring secure sourcing and vulnerability reporting. If a non-EU component manufacturer does not sell directly within the EU, the compliance burden shifts to the importer. Additionally, the CRA mandates the inclusion of a Software Bill of Materials (SBOM) as a crucial artifact for tracking vulnerabilities. Although digital services such as Software-as-a-Service (SaaS) fall outside the CRA's scope, the NIS2 Directive complements it by imposing cybersecurity and incident reporting obligations on critical service providers; for instance, cellular network operators.

Certain PDEs regulated under stringent sector-specific frameworks-such as automobiles and medical devices-are exempt from the CRA. Open Source Software (OSS) is similarly excluded, provided it is not monetized, as it does not constitute a commercial activity. Furthermore, official clarifications confirm that neither funding structures nor development conditions influence the commercial status of OSS[1].

New Product Liability Directive, 2024. The revised Product Liability Directive (PLD) [4] has replaced the 1985 directive, which originally established a no-fault strict liability framework within the European Union. This modernization adapts product liability regulations to the digital age, ensuring that individuals, including consumers, may seek compensation from manufacturers on a strict liability basis for defective products and, in certain cases, their components within the EU market. The primary objective of the new PLD is to streamline the claims process for damages arising from product defects.

The updated directive significantly broadens liability to encompass nearly all supply chain operators, thereby enhancing consumer protection irrespective of a product's origin, whether from within or outside the EU. Moreover, online marketplaces may also be held liable if they function as de facto sellers, though they can avoid such liability by providing details of the manufacturer's EU representative. Consumers are granted access to claim-related information while ensuring confidentiality, and they may seek compensation in complex cases, including those involving breaches of safety and security regulations such as the Cyber Resilience Act (CRA) or the AI Act. The elimination of arbitrary thresholds further guarantees full compensation for damages sustained.

The new PLD adopts a comprehensive definition of "product," encompassing physical goods, raw materials, and standalone software, including software integrated with artificial intelligence components. It explicitly addresses emerging technological concerns, including cybersecurity vulnerabilities, essential digital services, and software updates. As with the CRA, free and open-source software that falls outside the scope of commercial activity remains exempt. Additionally, the directive defines "component" broadly to include any integrated element, such as software libraries, raw materials, or services-such as software-initiated remote calls to a Software-as-a-Service (SaaS) instance-though standalone services generally remain outside its scope.

3.6.3 Challenges and Future Work

Based on Section 3.6.2, it seems that the strong-handed combination of ex ante safety and ex post strict liability regulations of the EU provides sufficient incentives for each supplier in the 5G supply chain. However, significant technical, legal, and incentive challenges still remain.

Technical challenges. Software products have been expanding in size and becoming more complex, in part from referencing and incorporating more third-party libraries and dependencies. Every additional library makes it more and more difficult to identify design, logic, and implementation vulnerabilities. Indeed, firms may well have no idea about existing or newly discovered vulnerabilities in the third-party code they use and deploy. For example, the widely discussed 2021 log4j incident left companies scrambling to assess whether their own systems are vulnerable owing to the popularity and large-scale reuse of the log4j package via the software supply chain [5]. The ineffectiveness of the lengthy assessment enabled attackers to continue exploiting the Log4Shell vulnerability throughout the next year. With nearly $30,000$ vulnerabilities disclosed in 2024, and increasing monotonically year over year, such incidents are bound to happen even more frequently [6]. Legal measures are therefore necessary, but not sufficient: developing and adopting a systematic approach and associated

[1] https://openforumeurope.org/eu-cyber-resilience-act-takes-a-leap-forward/

guidelines for exploring supply chain security risks is therefore a must. Furthermore, while the application of SBOMs could improve transparency, open-source and easy-to-use tools for SBOM creation, management, and analysis are sorely needed.

Another challenge is posed by the increasingly complex nature of 5G and 6G network ecosystems: the traditional perimeter-based cybersecurity controls deployed by telcos are no longer adequate. In fact, when components (nodes, ML models, software, etc.) of the network architecture themselves cannot be assumed trustworthy anymore, a Zero-Trust Architecture [12] is required. In such a security architecture, components need to be continuously monitored for anomalous behavior; how to achieve this without compromising performance and usability is far from trivial and an active research topic.

Legal and incentive challenges. Next to the technical challenges, we believe that better quality software (hardware, firmware, etc.) will not be produced without aligning economic incentives in their respective ecosystems. The most critical misalignment is that the harm caused by software problems is, by and large, shouldered by consumers, not vendors. This lack of liability means software vendors have every incentive to rush low-quality software onto the market and no incentive to enhance quality control. While European legislation, including the CRA, the PLD, and also NIS2 [2] from the 5G network operator side, defines strong incentives via penalties for non-compliance, in other parts of the world, the US, in particular, may require a predominantly market-based approach with minimal legislative elements. Such a complex mechanism is still very much under research currently, although strict requirements for information systems sold to federal institutions have already been passed [7]; these may have a beneficial spillover effect on the telecommunications sector. Furthermore, a promising proposal based on a mandatory but minimal product-based audit (floor) and a voluntary but liability-waiving process-based audit (ceiling) has been put forward to align the incentives of all stakeholders [13].

3.6.4 References

[1] M. Lyu, J. Farooq, and Q. Zhu, "Mapping Cyber Threats in the 5G Supply Chain: Landscape, Vulnerabilities, and Risk Management," *IEEE Network*, vol. 39, no. 1, pp. 251–260, 2025. doi: 10.1109/MNET.2024.3439011.

[2] THE EUROPEAN PARLIAMENT AND THE COUNCIL 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..." Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2555 (accessed 1 Feb. 2025).

[3] THE EUROPEAN PARLIAMENT AND THE COUNCIL 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)," Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R2847 (accessed 1 Feb. 2025).

[4] THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION, "Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC ," Available: https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng (accessed 1 Feb. 2025).

[5] R. Hiesgen, M. Nawrocki, T. C. Schmidt, and M. Wählisch, “The Log4j Incident: A Comprehensive Measurement Study of a Critical Vulnerability,” *IEEE Trans. Network and Service Management*, vol. 21, no. 6, pp. 5921–5934, 2024. doi: 10.1109/TNSM.2024.3440188.

[6] MITRE, “CVE Metrics.” Available: https://www.cve.org/about/Metrics (2025).

[7] White House, “Executive Order on Strengthening and Promoting Innovation in the Nation’s Cybersecurity,” Available: https://www.federalregister.gov/documents/2025/01/17/2025-01470/strengthening-and-promoting-innovation-in-the-nations-cybersecurity (Jan. 16, 2025).

[8] Federal Office for Information Security, Germany, “Technical Guideline TR-03183: Cyber Resilience Requirements for Manufacturers and Products,” Available: https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Technische-Richtlinien/TR-nach-Thema-sortiert/tr03183/tr-03183.html?nn=132646 (accessed 1 Feb. 2025).

[9] M. Ozger, I. Godor, A. Nordlow, T. Heyn, S. Pandi, I. Peterson, A. Viseras, J. Holis, C. Raffelsberger, A. Kercek, et al., “6G for connected sky: A vision for integrating terrestrial and non-terrestrial networks,” in *Proc. 2023 Joint European Conf. Networks and Communications & 6G Summit (EuCNC/6G Summit)*, pp. 711–716, 2023. Available: https://arxiv.org/abs/2305.04271

[10] M. A. Uusitalo, P. Rugeland, M. R. Boldi, E. C. Strinati, P. Demestichas, M. Ericson, G. P. Fettweis, M. C. Filippou, A. Gati, M.-H. Hamon, et al., “6G vision, value, use cases and technologies from European 6G flagship project Hexa-X,” *IEEE Access*, vol. 9, pp. 160004–160020, 2021. Available: https://doi.org/10.1109/ACCESS.2021.3130030

[11] N. Papernot, P. McDaniel, A. Sinha, and M. P. Wellman, “Sok: Security and privacy in machine learning,” in *Proc. 2018 IEEE European Symp. Security and Privacy (EuroS&P)*, pp. 399–414, 2018. Available: https://doi.org/10.1109/EuroSP.2018.00035

[12] Rose, Scott; Borchert, Oliver; Mitchell, Sean; and Connelly, Sean, “Zero Trust Architecture,” *NIST Special Publication 800-207*, National Institute of Standards and Technology, Gaithersburg, MD, USA, 2020. Available: https://doi.org/10.6028/NIST.SP.800-207

[13] G. Biczók, S. Romanosky, and M. Liu, “Realigning Incentives to Build Better Software: A Holistic Approach to Vendor Accountability,” *arXiv preprint arXiv:2504.07766*, 2025. Available: https://arxiv.org/abs/2504.07766

Chapter 4

Evolving Threats and Protective Strategies

Chapter Editors: An Braeken[1], Miranda Harizaj [2], Gergely Biczók [3], Gurjot Singh Gaba [4]

[1] Vrije Universiteit, Brussel, Belgium
[2] Polytechnic University of Tirana, Albania
[3] CrySyS Lab, Budapest University of Technology and Economics, Hungary
[4] Department of Computer and Information Science (IDA), Linköping University (LiU), Sweden

4.1 Introduction

Next-generation wireless architectures are characterized by a multi-dimensional and synchronously evolving attack surface, a critical aspect underscored by the diverse cases presented in this chapter. The rapid advancement of technologies like 5G-Advanced and 6G, coupled with the proliferation of interconnected devices and the integration of artificial intelligence, introduces unprecedented vulnerabilities that span from the foundational physical layer to complex application-specific contexts and even human factors. This chapter delves into the dynamic landscape of security challenges and the protective measures being developed to counter them. We begin by exploring the fundamental security protocols for future wireless communication systems, examining their innovations and the inherent implementation challenges. Following this, we investigate specific technical threat vectors, focusing on advanced fingerprinting techniques, including website fingerprinting in anonymous networks and RF fingerprinting for wireless device authentication, and the critical implications of Wi-Fi signal spoofing in localization.

Recognizing that technology alone cannot guarantee security, we then shift our focus to the human-centered aspects of cybersecurity, emphasizing the role of user behavior and awareness. Finally, the chapter concludes by examining security challenges and solutions within specific high-stakes application contexts, including the Internet of Medical Things (IoMT) and wireless network security in aviation, highlighting the unique threats and protective measures required for these critical verticals. By systematically analyzing these evolving threats and the corresponding protective strategies, this chapter aims to provide a comprehensive understanding of the current security landscape and future research directions for robust next-generation wireless systems.

4.2 Security Protocols for Future Wireless Communication Systems: Innovations and Implementation Challenges

Authors: Gurjot Singh Gaba[1], Andrei Gurtov[1], Evangelia Konstantopoulou[2], Nicolas Sklavos[2], An Braeken[3], Pardeep Kumar[4], Madhusanka Liyanage[5], Kapal Dev[6]

[1] Department of Computer and Information Science (IDA), Linköping University, Sweden
[2] SCYTALE Group, Computer Engineering and Informatics Department, University of Patras, Hellas
[3] Department of Engineering Technology (INDI), Vrije Universiteit, Belgium
[4] Department of Computer Science, Warwick Manufacturing Group, Warwick University, UK
[5] School of Computer Science, University College Dublin, Ireland
[6] Department of Computer Science, Munster Technological University, Ireland

4.2.1 Introduction

The security landscape of future wireless communication systems is becoming increasingly complex due to the growing scale of connected devices, advancements in cryptographic threats, and evolving network architectures. A major concern is the rise of quantum computing, which threatens traditional public-key encryption methods like RSA and ECC, potentially rendering them obsolete. While symmetric encryption schemes such as AES remain secure, they may require increased key sizes to counteract quantum threats [1].

Additionally, the proliferation of connected devices, ranging from high-performance edge computing nodes to low-power IoT sensors, necessitates scalable and efficient security solutions. Traditional security protocols often struggle to meet these demands, calling for new approaches that provide both resilience and efficiency. Emerging technologies, including post-quantum cryptography, AI-driven security mechanisms, and decentralized authentication frameworks, are being explored to mitigate these challenges and ensure secure, scalable, and adaptive wireless security solutions.

4.2.2 State of the Art

Recent advances in wireless network security have led to the exploration and adoption of several innovative technologies and approaches designed to address emerging threats, particularly in the context of quantum computing, resource-constrained environments, and decentralized architectures.

Post-quantum cryptographic algorithms, such as lattice-based and code-based encryption schemes, have emerged as viable solutions to counter the threats posed by quantum computing. Unlike traditional cryptographic methods that rely on hard problems like integer factorization and discrete logarithms, post-quantum techniques offer resistance against quantum attacks while preserving robust security properties [2]. However, these schemes often come with increased communication overhead rather than computational complexity, which can challenge their deployment in environments with limited bandwidth or energy resources. To address uncertainties in current standardization efforts and provide a path for gradual integration, hybrid cryptographic approaches that combine classical and post-quantum mechanisms have been proposed. These offer enhanced security guarantees during the transition phase and maintain long-term resilience across diverse devices and networks [3].

In parallel, the proliferation of Internet of Things (IoT) devices and edge computing systems has driven the need for lightweight cryptographic solutions tailored to resource-constrained environments. Lightweight protocols, primarily based on symmetric key primitives [4, 5], have been developed to minimize computational burden, memory usage, and power consumption. While symmetric encryption is highly efficient, it suffers from scalability and key management limitations, as each device pair must share a unique secret key. This introduces vulnerabilities such as exposure risk if a key is compromised. Furthermore, symmetric schemes lack built-in mechanisms for non-repudiation. Nevertheless, due to their performance benefits, they remain essential for real-time applications, bulk encryption tasks, and scenarios where low-latency operation is critical [6].

The integration of artificial intelligence (AI) and machine learning (ML) into wireless security systems has further transformed the threat detection landscape. AI-driven security mechanisms enable real-time analysis of network behavior, facilitating dynamic and adaptive threat detection. These systems learn from evolving attack patterns, allowing them to respond to previously unseen threats. This capability makes AI particularly effective in complex and changing wireless environments. However, the effectiveness of such systems is contingent upon access to large, diverse datasets and substantial computational resources for training and operation, which may limit their applicability in constrained settings.

Another innovative direction involves the application of blockchain technology to enhance trust and authentication in wireless systems. Blockchain's decentralized and tamper-resistant ledger provides a robust foundation for secure identity management and device authentication without reliance on centralized authorities [7]. This is especially beneficial in IoT networks, where establishing trust among numerous heterogeneous devices is a significant challenge. Despite its advantages, blockchain implementations can be resource-intensive, with high energy consumption and computational demands that may affect scalability and responsiveness.

Complementing these higher-layer techniques, physical layer security (PLS) has gained traction as a method to safeguard data transmissions by leveraging the inherent properties of the wireless channel [8]. Techniques such as channel-based encryption use the randomness and reciprocity of wireless signals to generate secure keys or provide confidential communication with minimal computational requirements [9]. These methods are particularly attractive for scenarios where low overhead is necessary. However, the effectiveness of PLS can be heavily influenced by environmental dynamics, such as mobility and signal interference, which limit its reliability and consistency in large-scale or rapidly changing networks.

To holistically evaluate these emerging technologies, a comparative analysis is presented in Table 4.1, outlining the trade-offs and suitability of each approach within wireless systems. This analysis underscores the importance of adopting a layered security strategy that combines multiple mechanisms to address the diverse and evolving challenges in securing future wireless networks.

Table 4.1: Comparison of Traditional and Emerging Security Protocols for Wireless Systems

Protocol Type	**Examples**	**Strengths**	**Limitations**	**Suitability for Future Systems**
Traditional	WPA2, IPSec	Strong encryption, widely implemented	Vulnerable to quantum attacks, scalability issues	Limited, needs upgrade
Post-Quantum Cryptography	Lattice-based, Code-based	Quantum-resistant, future-proof	Computationally intensive, standardization needed	High, essential for 6G
AI-Driven Security	Machine learning, anomaly detection	Real-time threat detection, adaptive	Requires training data, computational resources	High, suitable for dynamic networks
Blockchain-Based	Decentralized authentication	Secure, tamper-proof, trust management	High energy consumption, complexity	Moderate, potential for IoT
Physical Layer Security	Channel-based encryption	Exploits wireless properties, low overhead	Limited by environment, less mature	High, complements other methods

4.2.3 Challenges and Future Work

Standardization and Interoperability A major challenge in implementing emerging security solutions is the need for global standardization to ensure interoperability across different devices and network infrastructures. Organizations such as IEEE and 3GPP are actively working on defining security benchmarks for next-generation wireless systems. However, achieving widespread adoption requires collaboration between industry, academia, and government bodies.

Integration with Existing Infrastructure Deploying new security protocols without disrupting existing wireless infrastructure presents a significant challenge. Many current networks are built on legacy systems that may not be compatible with advanced cryptographic methods. Transitioning to post-quantum cryptography, for example, requires careful planning to minimize disruptions while maintaining robust security.

Computational Overhead and Scalability Advanced security mechanisms, such as post-quantum cryptography and AI-driven threat detection, often require high computational resources and high communication costs. This poses challenges for large-scale deployment, particularly in IoT environments with constrained devices. Future research should focus on optimizing these methods to enhance efficiency while maintaining strong security guarantees.

Implementation Aspects and Performance Security mechanisms are often deployed in embedded systems and handheld devices, such as mobile equipment, smartphones, and e-health gadgets. The efficiency of these security implementations depends on factors such as processing power, memory constraints, energy consumption, and system latency. In highly constrained environments, including the IoT and next-generation wireless networks (5G/6G), security solutions must operate within strict technical limitations while ensuring minimal impact on system performance. This requires careful optimization of cryptographic computations, authentication protocols, and data integrity mechanisms to balance security with system efficiency.

Real-World Testing and Deployment Real-world testing of advanced security solutions, especially those targeting quantum threats, is challenging due to the limited availability of quantum computers. Unlike traditional security protocols that can be tested on existing infrastructure, validating post-quantum cryptographic approaches requires access to high-performance quantum machines, which are currently restricted to a few organizations. This limitation makes it difficult to assess their real-world viability under genuine quantum attack scenarios. As a result, researchers rely on simulations and theoretical models, which, while insightful, do not fully replicate quantum threats. Expanding access to quantum computing and developing standardized testing frameworks will be essential for evaluating the practicality of these emerging security solutions.

4.2.4 Conclusion

Future wireless security must address challenges related to quantum threats, large-scale connectivity, and dynamic network environments. Emerging solutions such as post-quantum cryptography, AI-driven security, blockchain-based authentication, and physical-layer security offer promising advancements. However, their implementation requires careful consideration of standardization, efficiency, and real-world applicability [9]. Striking a balance between security requirements and key implementation factors, such as performance, resource allocation, and energy efficiency, is crucial for next-generation systems. A proactive, multi-layered security approach will be essential to ensuring resilience, scalability, and long-term protection in future wireless networks.

4.2.5 References

[1] Jaques, Samuel, “Quantum Attacks on AES: When Do We Need to Worry About a Structureless, Quantum, Known Plaintext Attack Against AES?” CHES 2024. Available: https://ches.iacr.org/2024/Jaques_CHES_2024.pdf

[2] Khan, Suleman; Gaba, Gurjot Singh; Gurtov, Andrei; Jansen, Leonardus J. A.; Mäurer, Nils; and Schmitt, Corinna, “Post-Quantum Secure Handover Mechanism for Next-Generation Aviation Communication Networks,” *IEEE Transactions on Green Communications and Networking*, vol. 8, no. 3, pp. 939–955, Sep. 2024. Available: https://doi.org/10.1109/TGCN.2024.3417298

[3] Braeken, A., “Flexible Hybrid Post-Quantum Bidirectional Multi-Factor Authentication and Key Agreement Framework Using ECC and KEM,” *Future Generation Computer Systems*, vol. 166, 107634, May 2025. Available: https://doi.org/10.1016/j.future.2024.107634

[4] Vandervelden, Thibaut; De Smet, Ruben; Steenhaut, Kris; and Braeken, An, “Symmetric-Key-Based Authentication Among the Nodes in a Wireless Sensor and Actuator Network,” *Sensors*, vol. 22, no. 4, 1403, Feb. 2022. Available: https://doi.org/10.3390/s22041403

[5] Konstantopoulou, Evangelia; Athanasiou, George S.; and Sklavos, Nicolas, “Review and Analysis of FPGA and ASIC Implementations of NIST Lightweight Cryptography Finalists,” *ACM Computing Surveys*, vol. 57, no. 10, pp. 1–35, Oct. 2025. Available: https://doi.org/10.1145/3721122

[6] Braeken, An, “Public Key Versus Symmetric Key Cryptography in Client–Server Authentication Protocols,” *International Journal of Information Security*, vol. 21, no. 1, pp. 103–114, Feb. 2022. Available: https://doi.org/10.1007/s10207-021-00543-w

[7] Ratnayake, R.; Liyanage, M.; and Murphy, L., “Evaluating Data Trust in Blockchain-Based IoT Systems Using Machine Learning Techniques,” ResearchGate Preprint. Available: https://www.researchgate.net/publication/384572427

[8] De Alwis, Chamitha; Kumar, Pardeep; Pham, Quoc-Viet; Dev, Kapal; Kalla, Anshuman; Liyanage, Madhusanka; and Hwang, Won-Joo, “Towards 6G: Key Technological Directions,” *ICT Express*, vol. 9, no. 4, pp. 525–533, Oct. 2022. Available: https://doi.org/10.1016/j.icte.2022.10.005

[9] Sklavos, Nicolas (Ed.), *Hardware Security and Trust: Design and Deployment of Integrated Circuits in a Threatened Environment*. Cham, Switzerland: Springer, Jan. 2017. Available: https://doi.org/10.1007/978-3-319-44318-8

4.3 Website Fingerprinting in Anonymous Networks: a Theoretical Lens on Trends, Challenges, and Research Frontiers

Authors: Arbena Musa[1], Blerim Rexha[1]

[1] Faculty of Electrical and Computer Engineering, University of Prishtina, Kosovo

4.3.1 Introduction

Anonymous communication networks (ACNs) such as Tor, I2P, Freenet, and Lokinet have become foundational infrastructures for ensuring online privacy. These systems function by concealing users' identities and routing traffic through layers of encryption to obscure source-destination relationships. Yet, even within these protective environments, adversaries can exploit side-channel information through a technique known as website fingerprinting (WF)-a form of traffic analysis that uses patterns in packet size, direction, and timing to infer which websites a user visits.

Anonymity networks are essential for shielding individuals from intrusive surveillance and fostering digital freedom. However, the continuous advancement of surveillance technologies and deanonymization techniques increasingly challenges the effectiveness of these networks. This evolving tension reflects a broader societal issue: the urgent need to cultivate a digital environment where individuals can interact, seek information, and communicate without sacrificing their right to privacy. Addressing this challenge demands not only technical innovation but also a nuanced understanding of the complex interplay between anonymity, visibility, and the evolving threats to both personal and collective privacy.

Figure 4.1 illustrates various digital fingerprinting techniques, highlighting their data collection methods, tracking scope, and privacy implications. These techniques range from device and browser fingerprints to behavioral and cross-device identifiers, forming a critical intersection between anonymity and identification in cybersecurity.

The growing capabilities of WF attacks have been significantly amplified by recent advances in machine learning, deep learning, and large-scale traffic analysis. These developments have enabled more accurate and generalizable attacks, often surpassing traditional limitations. The problem is especially acute in wireless network contexts-ranging from public Wi-Fi to mobile and ad-hoc networks-where traffic variability, device diversity, and exposure to local adversaries increase users' vulnerability to passive monitoring.

Here, we provide a conceptual overview of the current state of website fingerprinting research, classifying attack methodologies and defense strategies while highlighting key trends and open challenges.

4.3.2 State of the Art

The body of WF research can be interpreted through three interwoven lenses: the epistemological orientation of the studies, the adversarial interaction paradigm and the network or application context. Together, these lenses help clarify both the diversity of current approaches and the underlying assumptions that shape them.

From an epistemological point of view, WF studies range from empirical to theoretical. Experimental work typically evaluates the performance of attacks or defenses on benchmark datasets under controlled conditions. These studies aim to demonstrate effectiveness in practical scenarios, measuring metrics such as classification accuracy or evasion success.

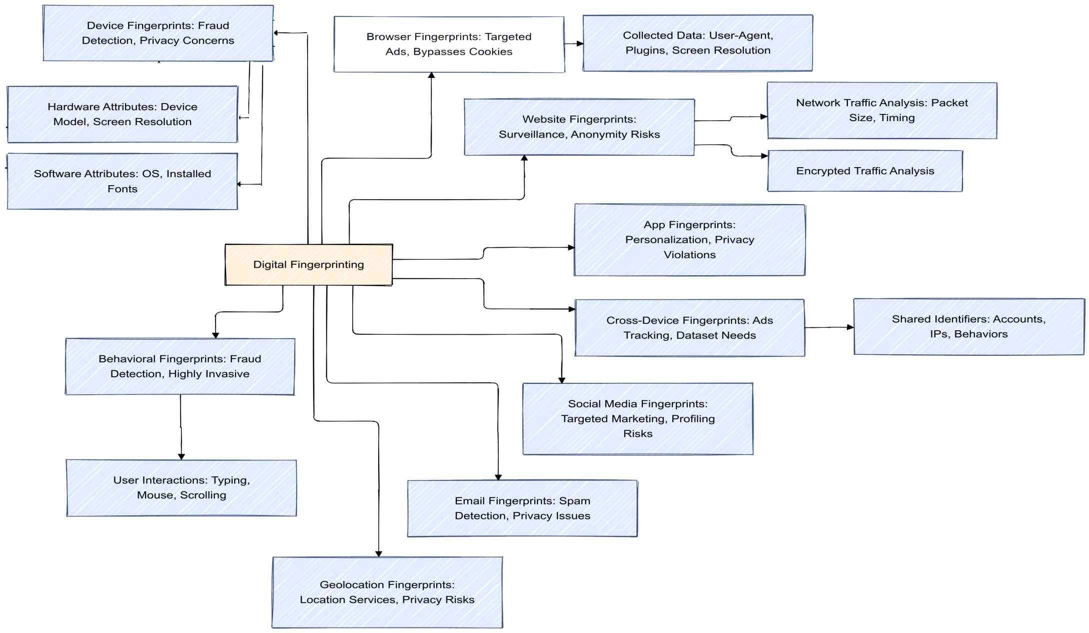

Device Fingerprints: Fraud Detection, Privacy Concerns
Browser Fingerprints: Targeted Ads, Bypasses Cookies
Collected Data: User-Agent, Plugins, Screen Resolution
Hardware Attributes: Device Model, Screen Resolution
Website Fingerprints: Surveillance, Anonymity Risks
Network Traffic Analysis: Packet Size, Timing
Software Attributes: OS, Installed Fonts
Encrypted Traffic Analysis
App Fingerprints: Personalization, Privacy Violations
Digital Fingerprinting
Cross-Device Fingerprints: Ads Tracking, Dataset Needs
Shared Identifiers: Accounts, IPs, Behaviors
Behavioral Fingerprints: Fraud Detection, Highly Invasive
Social Media Fingerprints: Targeted Marketing, Profiling Risks
User Interactions: Typing, Mouse, Scrolling
Email Fingerprints: Spam Detection, Privacy Issues
Geolocation Fingerprints: Location Services, Privacy Risks

Figure 4.1: Extensive Overview of Digital Traces [1].

In contrast, theoretical work delves into formal models, privacy limits, and abstract threat frameworks, seeking to generalize findings and inform principled system design.

A second axis of classification involves the nature of the adversarial strategy. Passive attacks dominate the literature, relying only on observation to classify web traffic based on statistical regularities or machine-learned patterns. These models often exploit burst sequences, packet timing, and flow directions to construct traffic fingerprints. Active attacks, though less common, introduce controlled perturbations, such as delayed injections or packet flows, to amplify signal differences. Although potentially more effective, active methods raise ethical questions and risk detection.

The third dimension concerns the system context and target scope. WF attacks have evolved beyond single-tab website classification to include more complex settings: multi-tab browsing, mobile users, hidden services, and cross-network scenarios. Many recent studies evaluate attacks on closed-world datasets with a limited set of monitored sites. Others adopt open-world settings that better simulate realistic usage but pose greater challenges due to class imbalance and noise. The inclusion of real-world wireless traffic remains limited, despite its growing relevance.

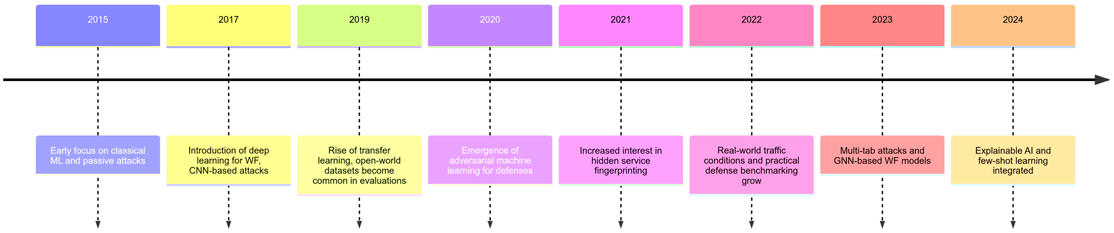

2015
2017
2019
2020
2021
2022
2023
2024
Early focus on classical ML and passive attacks
Introduction of deep learning for WF, CNN-based attacks
Rise of transfer learning, open-world datasets become common in evaluations
Emergence of adversarial machine learning for defenses
Increased interest in hidden service fingerprinting
Real-world traffic conditions and practical defense benchmarking grow
Multi-tab attacks and GNN-based WF models
Explainable AI and few-shot learning integrated

Figure 4.2: Key milestones in website fingerprinting research across the last decade.

A review of recent literature reveals several methodological shifts, many of which are reflected in Figure 4.2 [2, 3]. Deep learning, particularly using convolutional neural networks (CNNs), transformers, and graph neural networks (GNNs) [4], has largely supplanted traditional classifiers. Transfer learning and few-shot learning have improved model adaptability, allowing attacks to generalize across different traffic environments [5]. Furthermore, adversarial machine learning is increasingly used both to craft robust attacks and to generate evasive perturbations as part of defense mechanisms [6]. These trends indicate a maturing field in which technical sophistication is balanced with growing concerns about generalization, scalability, and ethics.

4.3.3 Challenges and Future Work

Despite notable advances, website fingerprinting remains a deeply contested space, with fundamental challenges yet unresolved. Wireless environments, in particular, present a set of unique difficulties that require fresh approaches and interdisciplinary collaboration.

A primary challenge is building resilience in unstable conditions. Network traffic is inherently variable-subject to jitter, intermittent connections, and bandwidth fluctuations. Most WF models assume relatively stable network behavior and degrade when faced with dynamic conditions. Future research must explore adaptive and noise-tolerant models that maintain performance without frequent retraining.

Another pressing concern is the development of lightweight defenses suitable for mobile and embedded systems. Many current techniques-such as traffic padding, morphing, or adversarial noise injection-incur substantial computational or bandwidth costs. These are infeasible for low-power devices or high-latency applications. defenses must evolve to become context-aware, responsive to traffic state, and efficient in both energy and data use.

Cross-domain generalization also remains elusive. Attacks that perform well in controlled settings often fail when deployed across different browsers, operating systems, or network types. Research into domain-invariant representations and robust training paradigms could close this gap, enhancing both attack realism and defense reliability.

In summary, as attacks become more adaptable and environments more complex, the challenge is to develop holistic, scalable, and ethically grounded approaches to defending anonymity-especially in the wireless age.

4.3.4 References

[1] Musa, Arbena; Vishi, Kamer; Martiri, Edlira; and Rexha, Blerim, "Our Digital Traces in Cybersecurity: Bridging the Gap Between Anonymity and Identification," *IEEE Access*, vol. 13, pp. 46909–46924, 2025. Available: https://doi.org/10.1109/ACCESS.2025.3551095

[2] Pan, Tianyao; Tang, Zejia; and Xu, Dawei, "A Practical Website Fingerprinting Attack via CNN-Based Transfer Learning," *Mathematics*, vol. 11, no. 19, 4078, Sep. 2023. Available: https://doi.org/10.3390/math11194078

[3] Liu, Peidong; He, Longtao; and Li, Zhoujun, "A Survey on Deep Learning for Website Fingerprinting Attacks and Defenses," *IEEE Access*, vol. 11, pp. 26033–26047, 2023. Available: https://doi.org/10.1109/ACCESS.2023.3253559

[4] Tan, Xiaobin; Peng, Chuang; Xie, Peng; Wang, Hao; Li, Mengxiang; Chen, Shuangwu; and Zou, Cliff, "Inter-Flow Spatio-Temporal Correlation Analysis Based Website Fingerprinting Using Graph Neural Network," *IEEE Transactions on Information Forensics and*

Security, vol. 19, pp. 7619–7632, 2024. Available: https://doi.org/10.1109/TIFS.2024.3441935

[5] Luo, Chenxiang; Tang, Wenyi; Wang, Qixu; and Zheng, Danyang, “Few-Shot Website Fingerprinting With Distribution Calibration,” *IEEE Transactions on Dependable and Secure Computing*, vol. 22, no. 1, pp. 632–648, Jan. 2025. Available: https://doi.org/10.1109/TDSC.2024.3411014

[6] Wang, Di; Zhu, Yuefei; Fei, Jin-Long; and Guo, Maohua, “CMAES-WFD: Adversarial Website Fingerprinting Defense Based on Covariance Matrix Adaptation Evolution Strategy,” *Computers, Materials & Continua*, vol. 79, no. 2, pp. 2253–2276, 2024. Available: https://doi.org/10.32604/cmc.2024.049504

4.4 Advancements and Challenges in RF Fingerprinting for Wireless Device Authentication in Next Generation Networks

Authors: Miranda Harizaj[1], Iraklis Symeonidis[2], Ali Kara[3]

[1] Faculty of Electrical Engineering, Polytechnic University of Tirana, Albania
[2] Department of Industrial Systems, RISE Research Institutes of Sweden
[3] Department of Electrical and Electronics Engineering, Faculty of Engineering, Gazi University, Ankara, Türkiye

4.4.1 Introduction

Next generation networks (beyond 5G) are expected to support ultra dense device connectivity with the proliferation of radio devices along with increased use of Internet of Things (IoT). Authentication of IoT devices typically rely on cryptographic algorithms for key exchange and unique device identifiers, typically software addresses, that is, MAC address. Key generation and exchange require relatively long time and complex computations which IoT devices mostly cannot afford. As RFF relies on unique characteristics of IoT transmitters, therefore, it could be enabling technology, and has been emerged as a candidate authentication technique for low power, computation-constrained devices of next generation networks. Radio frequency fingerprinting (RFF) is a signal intelligence method whereby unique characteristics of radio transmitting devices are extracted to aid device identification. Unique characteristics, so-called fingerprints or signatures, specifically, are due to the imperfections in manufacturing of components or chipsets like amplifiers, filters and clock generators. These imperfections lead to variations in the signal characteristics that result in, for example, phase offset, IQ imbalance, clock skew, and some others [1].

As fingerprints play a unique role in the identification of RFF devices, fingerprint extraction is the most critical stage of RFF techniques. Human-made or human-extracted features are based on manually extracted hardware features, where experts are expected to work on the received signal characteristics to identify the most robust features. However, the recent trend is to employ artificial intelligence (AI) to extract features along with classification. It may also be possible to employ hybrid, that is, combining human-extracted features with artificial intelligence, that is, integrating deep learning techniques [2]. Despite recent advancements, there remains a gap in practical, scalable deployment of RFF systems, particularly concerning lightweight AI model optimization, adaptability to real-world environmental variables, and robustness in open-set scenarios [3].

4.4.2 State of the Art

Recent research efforts have been directed toward AI-aided RFF identification systems.A comprehensive overview of RFF methods, covering both traditional and AI-driven approaches while identifying key scalability challenges, is provided in [1]. The findings of experimental studies conducted with open-access datasets suggest that AI-aided RFF techniques have been a standard approach [3]. The targeted applications along with operational bands and protocol are quite diverse. LoRA (868 MHz), ZigBee (2505MHz), Bluetooth(2400 MHz), ADS-B (1090 MHz) and WiFi (2400 MHz) are some common protocols [4]. Smart city ecosystem seems to be the major application domain as huge number of IoT devices are to be deployed in this ecosystem. To be specific on AI-aided RFF, recent advancements in Radio Frequency

Fingerprinting (RFF) have increasingly leveraged deep learning techniques to strengthen device authentication, particularly within IoT networks. Additionally, the practical deployment challenges and potential of integrating RFF systems into industrial infrastructures such as smart grids as an integral part of smart city ecosystem, addressing regulatory, implementation, and scalability aspects crucial for real-world industrial environments is emphasized in [5]. Further, the detailed RFF identification process and its practical limitations, such as low signal-to-noise ratios and the need for robust open-set recognition are outlined in [6]. Several studies have explored lightweight deep learning models for LoRa devices, focusing on balancing classification accuracy with real-time deployment feasibility [7].

4.4.3 Challenges and Future Work

The challenges and future works can be directed to the following major aspects of RFF identification systems.

Open-set authentication Contemporary RFF aims to authenticate devices in an open-set scenario, where rogue devices attempt to access network resources. Recent advances in AI-aided RFF techniques have enabled the extraction of highly generalized features, crucial for detecting rogue devices not encountered during the training phase and enhancing authentication. Open-set authentication, that is, the capability of the system to recognize and reject devices that were not seen during training is very critical.

Development and Optimization of AI Models for Implementation on Edge Devices Performance metrics such as classification accuracy and openness are widely recognized as crucial in RFF research efforts. Low inference latency holds significant importance, particularly in the context of next generation networks where ultra-reliable and low-latency communications (URLLC) is paramount. Practical IoT deployments may require bi-directional communication necessitating the implementation of RFF directly on edge devices. Then, there is a need to develop lightweight AI models optimized for edge deployment. These models should authenticate and classify devices in real time while ensuring that classification accuracy is upheld even at a high value of openness.

Characterization and Mitigation of Operational Effects on Performance Environmental and operational conditions, the physical layer or the channel, significantly impact classification accuracy. This is an inherent challenge in RFF as it operates at the physical layer. Aside from wireless channel effects which has been studied to some extent, temperature variations between training and inference stages may reduce classification accuracy. It is pertinent to note that the effect of aging has not been characterized over a long duration of time in a detailed study. Moreover, it is reasonable to assume that an RFF receiver used in the training and testing stage shall be different from the one deployed in a practical application. The studies have shown that changing the receiver between training and inference significantly compromises classification accuracy.

Testing and validation in an application domain It is quite important to test and validate RFF authentication and access control system in a practical setting, typically, one or two components of a smart city ecosystem. For example, smart grids are vital to energy infrastructure, and the cybersecurity of IoT devices is paramount. Such demonstration validates the RFF system under real-world conditions, assessing its performance in diverse environmental and operational scenarios. It may also demonstrate the RFF authentication capability in a complex and dynamic environment, ensuring robustness and reliability.

RFF: A Double-Edged Sword for Security and Privacy Radio Frequency Fingerprinting (RFF) has a dual nature: it can serve as a powerful tool for enhancing security, authentication, and anti-fraud measures, yet it also poses significant threats to user privacy if misused. Its ability to uniquely identify devices based on hardware-level signal imperfections makes it highly effective for device-level authentication, access control, and even theft prevention-particularly within IoT ecosystems. However, this same persistent identifiability raises serious concerns [9, 8]. RFF enables persistent, passive, and unconsented tracking and profiling of users across networks and locations, often without their knowledge-potentially violating privacy regulations such as the GDPR [10] in EU. Unlike changeable identifiers such as IP or MAC addresses, RF fingerprints are difficult to spoof or reset, creating opportunities for long-term behavioral surveillance. These fingerprints can be used to infer patterns in movement, device usage, and other behaviors, especially when combined with auxiliary data sources. Malicious actors, overreaching institutions, or even legitimate entities may exploit these identifiers for targeted surveillance and profiling, with particularly serious implications in authoritarian contexts. The lack of user control and the potential for data leakage further amplify these vulnerabilities, particularly when RF data is intercepted or used to link devices to personal information. Therefore, while RFF can enhance user security, its deployment must be guided by strong privacy-by-design principles, robust technical safeguards, and meaningful user consent to ensure it protects rather than undermines individual rights.

4.4.4 References

[1] A. Jagannath, J. Jagannath, and P. S. K. V. Pattanshetty, “A Comprehensive Survey on Radio Frequency (RF) Fingerprinting: Traditional Approaches, Deep Learning, and Open Challenges,” *Computer Networks*, vol. 219, Art. no. 109455, Dec. 2022. doi: 10.1016/j.comnet.2022.109455.

[2] Zhang, Junqing; Shen, Guanxiong; Saad, Walid; and Chowdhury, Kaushik, “Radio Frequency Fingerprint Identification for Device Authentication in the Internet of Things,” *IEEE Communications Magazine*, vol. 61, no. 10, pp. 110–115, Oct. 2023. Available: https://doi.org/10.1109/MCOM.003.2200974

[3] Demiroğlu, Harun Şenol; Awan, Maaz Ali; and Kara, Ali, “An Overview of Challenges to Long-Term Sustainability and Scalability of Radio Frequency Fingerprinting,” in *Proc. 2024 6th Int. Conf. on Communications, Signal Processing, and their Applications (ICCSPA)*, IEEE, July 2024, pp. 1–6. Available: https://doi.org/10.1109/ICCSPA61559.2024.10794273

[4] Yan, Gaoli; Fu, Xue; Wang, Yu; Zhang, Qianyun; and Gui, Guan, “Radio Frequency Fingerprint Identification Towards Statistical and Deep Learning Features: Review, Recent Results and Future Directions,” *Peer-to-Peer Networking and Applications*, vol. 18, no. 3, May 2025. Available: https://doi.org/10.1007/s12083-024-01902-9

[5] Awan, Maaz Ali; Dalveren, Yaser; Catak, Ferhat Ozgur; and Kara, Ali, “Deployment and Implementation Aspects of Radio Frequency Fingerprinting in Cybersecurity of Smart Grids,” *Electronics*, vol. 12, no. 24, 4914, Dec. 2023. Available: https://doi.org/10.3390/electronics12244914

[6] A. Ahmed, B. Quoitin, A. Gros, and V. Moeyaert, “A Comprehensive Survey on Deep Learning-Based LoRa Radio Frequency Fingerprinting Identification,” *Sensors*, vol. 24, no. 13, Art. no. 4411, Jul. 2024. doi: 10.3390/s24134411.

[7] Ahmed, Aqeel; Quoitin, Bruno; Gros, Alexander; and Moeyaert, Veronique, “A Comprehensive Survey on Deep Learning-Based LoRa Radio Frequency Fingerprinting Identification,” *Sensors*, vol. 24, no. 13, 4411, July 2024. Available: https://doi.org/10.3390/s24134411

[8] M. Safi, S. Dadkhah, F. Shoeleh, H. Mahdikhani, H. Molyneaux, and A. A. Ghorbani, “A Survey on IoT Profiling, Fingerprinting, and Identification,” *ACM Transactions on Internet of Things*, vol. 3, no. 4, pp. 1–39, Nov. 2022. doi: 10.1145/3539736.

[9] Abbas, Sohail; Abu Talib, Manar; Nasir, Qassim; Idhis, Sally; Alaboudi, Mariam; and Mohamed, Ali, “Radio Frequency Fingerprinting Techniques for Device Identification: A Survey,” *International Journal of Information Security*, vol. 23, no. 2, pp. 1389–1427, Apr. 2024. Available: https://doi.org/10.1007/s10207-023-00801-z

[10] Council of the EU Final Compromised Resolution. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

4.5 Spoofing of Wi-Fi signals in fingerprinting-based localization

Authors: Juraj Machaj[1], Peter Brida[1]

[1] Faculty of Electrical Engineering and Information Technology, University of Zilina, Slovakia

4.5.1 Introduction

Wi-Fi-based indoor positioning has gained significant attention due to the ubiquitous availability of Wi-Fi infrastructure and its integration into consumer devices. However, since Wi-Fi operates in an unlicensed frequency spectrum, it is vulnerable to security threats such as Access Point (AP) spoofing, which can distort positioning accuracy. Therefore, it is important to address AP spoofing in fingerprinting-based localization.

4.5.2 State of the Art

AP spoofing involves malicious entities impersonating legitimate Wi-Fi access points, causing devices to receive signals from fake APs. This deception can result in higher localization errors, which can challenge location-based services. Several studies have explored the vulnerabilities of WLAN-based positioning systems to such spoofing attacks and proposed solutions to reduce the impact of spoofing signals.

It is assumed that AP spoofing can affect the performance of the localization system significantly, reducing the accuracy of the system and misleading the users and services that rely on estimated position. The impact of different numbers of spoofed APs on KNN (K-Nearest Neighbour) algorithm under two scenarios was evaluated using the UJIIndoorloc dataset in [1], and results are shown in Figure 4.3. The first scenario considered in this case is represented by randomly spoofed APs, with RSS values ranging between -30 dBm and -70 dBm. On the other hand, the second scenario was based on the replay of Received Signal Strength (RSS) samples collected at different positions, the RSS values in this case were between 0 dBm and -100 dBm. It should be noted here that since KNN localization is deterministic and is not trained on radiomap data the impact of spoofing is limited. However, the impact on machine learning based localization algorithms is more significant.

From the figure, it can be seen that the impact of the spoofing increases with the number of the spoofed APs which is expected behavior. Moreover, the impact was more significant in case of the RSS replay spoofing.

Yang et al., [3] developed a technique for the detection of false signals based on spatial correlation. Their method is based on the assumption that the RSS measured by the landmarks is correlated to the location of the transmitter and distance from the landmark. Therefore, it should be possible to detect spoofed localization requests. However, implementation requires additional infrastructure, and landmarks will also detect spoofed signals.

Furthermore, Jiang et al., [4] proposed a virtual MAC spoofing detection method based on deep learning algorithms and anomaly detection to analyze patterns in network traffic. The proposed solution has the ability to detect MAC spoofing attempts. The authors reported an average detection accuracy of 95 %. However, the solution requires Channel State Information (CSI) measurements, which are not readily available on consumer devices, and active communication with the device that has a spoofed MAC address.

Another notable approach was proposed by Restuccia et al., in [5]. The proposed Location Validation System (LVS) protects the location system service from spoofing attacks. The

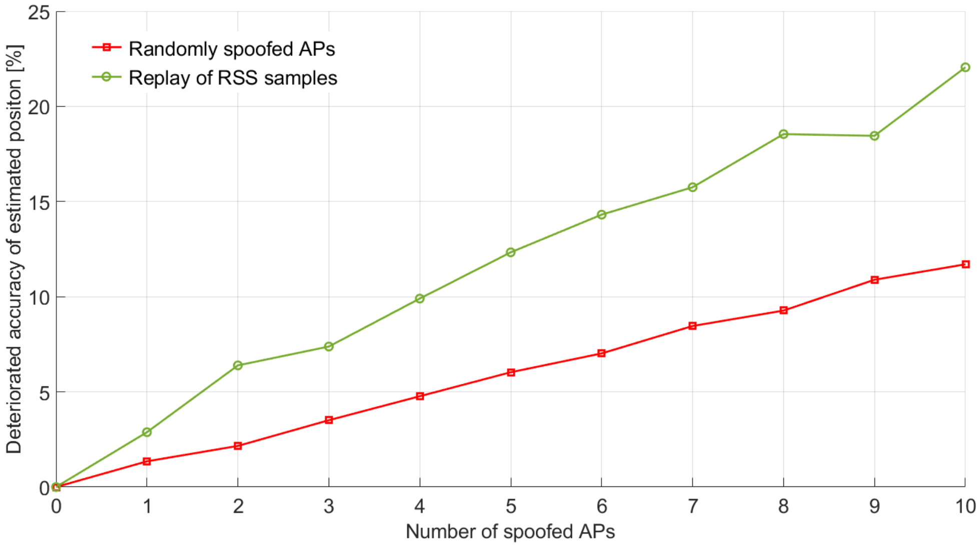


Figure 4.3: Impact of different number of spoofed APs on KNN algorithm [1].

solution is based on the validation of position by neighbouring nodes in validation rounds. These neighbouring nodes could be represented by mobile devices, e.g. smartphones, that can operate as Wi-Fi hotspots. The proposed solution thus can handle spoofed position estimates, however, if there are signals from fake APs in the area all devices may be affected by spoofing and thus this approach is not viable in this scenario.

In addition to software-based solutions, hardware-based fingerprinting approaches have been explored. Tian et al., [6] discussed the potential of identifying unique devices through their hardware characteristics, using the Wasserstein metric to detect spoofing attacks in Wi-Fi networks. The idea is based on the fact that each device has slightly different hardware characteristics, which could be used to detect signals generated by fake APs. In the proposed method frequency offsets of signals received from APs with different MAC addresses are compared, in order to validate that these signals are not transmitted by a single device. In order to perm the classification, a large number of collected samples is required.

However, the methodologies mentioned above require additional measurements of RSS samples or use CSI measurements, which are not readily available on consumer devices. The Spatially Filtered K-Nearest Neighbors (SFKNN) algorithm has been proposed to improve the detection of AP spoofing in Wi-Fi fingerprinting-based positioning systems [7]. SFKNN integrates spatial filtering with the traditional KNN approach, analyzing spatial inconsistencies in RSS data to identify potential spoofing activities. By incorporating spatial context, SFKNN aims to improve the robustness of positioning systems against malicious interference, effectively distinguishing between legitimate signal variations and those induced by spoofing attempts.

The overview of the above-mentioned solutions for AP spoofing detection is summarized in Table 4.2. From the table it is clear that spoofing detection based on metadata from radiomap implemented in SFKNN can provide results without a need for additional measurements. Moreover, the processing can be performed offline during the setup of the localization system thus the complexity of the system is not affected significantly. However, from the results presented in [7] it is clear that the algorithm has a relatively high number of false-positive results, which was around 10 %. Moreover, detection works better in the case of random spoofing compared to RSS replay. The spoofing detection was above 90 % and above 70 %

Table 4.2: Summary of methods for spoofing detection in Wi-Fi localization.

Ref.	Contribution	Limitation
[3]	Detection of spoofed information using spatial correlation.	Requires installation of landmarks in localization area, resulting in increased cost of both infrastructure and maintenance.
[5]	Detection of spoofed position estimates using data from neighbouring nodes	Requires cooperation between individual devices in the area. It does detect spoofed positions, not spoofed signals.
[4]	Use of CSI measurements and deep learning.	Limited access to CSI measurements on off-the-shelf devices, which are widely used in Wi-Fi localization
[6]	Detection of multiple signals transmitted by a single fake AP.	Requires a large number of samples, thus introducing a significant delay in the localisation process.
[7]	Detection based on metadata from radio map.	Preprocessing of data in the radio map during the offline phase.

when the number of spoofed APs was ≥ 3 for random and RSS replay spoofing, respectively.

4.5.3 Challenges and Future Work

Despite these advances, there are still challenges in effectively mitigating AP spoofing. The dynamic nature of indoor environments and the proliferation of Wi-Fi devices require a continuous refinement of the detection algorithms. Future research directions include integrating machine learning techniques for adaptive filtering and conducting extensive testing in real-world deployment scenarios to ensure robustness against evolving spoofing techniques. Moreover, solutions should be proposed to further improve the spoofing detection accuracy.

Acknowledgement

This work has been partially supported by the NATO SPS G6140 "Advanced technologies for Physical ResIlience Of cRitical Infrastructures (APRIORI)".

4.5.4 References

[1] J. Machaj, P. Brida, and B. Adamec, "Effect of Wi-Fi Access Points Spoofing on Fingerprinting Localization," in *Proc. 2024 34th Int. Conf. Radioelektronika (RADIOELEKTRONIKA)*, IEEE, Apr. 2024, pp. 1–5. doi: 10.1109/RADIOELEKTRONIKA61599.2024.10524072.

[2] J. Torres-Sospedra, R. Montoliu, A. Martinez-Uso, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "UJIIndoorLoc: A New Multi-Building and Multi-Floor Database for WLAN Fingerprint-Based Indoor Localization Problems," in *Proc. 2014 Int. Conf. Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, Oct. 2014, pp. 261–270. doi: 10.1109/IPIN.2014.7275492.

[3] J. Yang, Y. Chen, W. Trappe, and J. Cheng, "Detection and Localization of Multiple Spoofing Attackers in Wireless Networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 1, pp. 44–58, Jan. 2013. doi: 10.1109/TPDS.2012.104.

[4] P. Jiang, H. Wu, C. Wang, and C. Xin, “Virtual MAC Spoofing Detection Through Deep Learning,” in *Proc. 2018 IEEE Int. Conf. Communications (ICC)*, May 2018, pp. 1–6. doi: 10.1109/ICC.2018.8422830.

[5] F. Restuccia, A. Saracino, S. K. Das, and F. Martinelli, “LVS: A WiFi-Based System to Tackle Location Spoofing in Location-Based Services,” in *Proc. 2016 IEEE 17th Int. Symp. A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Jun. 2016, pp. 1–4. doi: 10.1109/WoWMoM.2016.7523533.

[6] Y. Tian, N. Zheng, X. Chen, and L. Gao, “Wasserstein Metric-Based Location Spoofing Attack Detection in WiFi Positioning Systems,” *Security and Communication Networks*, vol. 2021, Art. ID 8817569, Apr. 2021. doi: 10.1155/2021/8817569.

[7] J. Machaj, C. Safon, S. Matúška, and P. Brída, “Detection of Access Point Spoofing in the Wi-Fi Fingerprinting Based Positioning,” *Sensors*, vol. 24, no. 23, Art. 7624, Nov. 2024. doi: 10.3390/s24237624.

4.6 Towards human-centred cybersecurity

Authors: Asreen Rostami[1], Kaja Fjørtoft Ystgaard[2], Camille Sivelle[2], Kamil Koniuch[3], Katrien De Moor[2]

[1] RISE Research Institutes of Sweden, Sweden

[2] Department of Information Security and Communication Technology, Norwegian University of Science and Technology (NTNU), Norway

[3] AGH University of Science and Technology, Poland

4.6.1 Introduction

Next-generation wireless systems, including 5G and 6G networks, are reshaping the digital landscape. These infrastructures promise new possibilities such as ultra-low latency, high-speed connectivity, and support for massive device connectivity. Their capabilities are enabled by advances in technologies such as machine learning (ML) and artificial intelligence (AI). With these developments, cutting-edge applications-such as the Internet of Things (IoT), Extended Reality, cloud computing, pervasive sensing, and real-time remote services-can thrive. However, the increasing complexity and interconnectivity of these systems expand the surface of potential security vulnerabilities, many of which elude resolution through technical solutions alone [19]. Notably, these technologies do not exist in isolation. They are embedded within broader socio-technical systems and shaped by human actors operating in diverse roles and contexts. As such, an adequate understanding of cybersecurity must extend beyond purely technical considerations to engage with the human, social, and institutional dimensions of digital security. To date, however, most dominant cybersecurity perspectives characterise the human element as the "weakest link" [6], framing users primarily as risks to be mitigated. This reductive stance overlooks the situated complexity of human behaviour and fails to account for the ways in which individuals contribute to cybersecurity resilience. A growing body of research counters this narrative, positioning users not merely as passive subjects or points of failure, but as active participants in the configuration, interpretation, and even creative adaptation of security systems (see e.g., [22]).

These shifts in conceptualizations have significant implications for practice. Despite increasing recognition of users' agency, the majority of security strategies and interventions remain grounded in outdated assumptions about user behaviour. For instance, common strategies, such as security awareness campaigns and user training, often rest on the premise that increased knowledge directly translates into secure practices or that users possess the necessary cognitive and technical competencies to implement complex security protocols. In practice, however, users have frequently reported fatigue and frustration with dealing with burdensome security tasks, such as managing multiple passwords or navigating inflexible, repetitive authentication systems. These burdens are unevenly distributed and disproportionately affect individuals with limited technical literacy and, as a result, contribute to the existing digital inequalities.

Positioned within the wider cybersecurity discourse, this paper therefore articulates a human-centred perspective on cybersecurity, offering a critical analysis of prevailing approaches and outlining key challenges for a more inclusive and contextually grounded cybersecurity agenda.

4.6.2 State of the Art

Conceptualisation and definition of cybersecurity Cybersecurity is conventionally defined as the combination of technologies, processes, and practices designed to protect systems, networks, and data from unauthorised access, damage, or disruption [4]. According to ISO/IEC 2382-8, computer security entails safeguarding digital resources from both accidental and intentional threats. Within this framework, the user – whether conceived as an end-user, an organisation, or members of the broader public who interact with or are impacted by technology – is typically framed as a behavioural subject whose compliance is to be shaped through training, policy enforcement, and system design.

Human-centred cybersecurity challenges this framing by adopting a broader, more integrated, and inclusive conceptualisation. It attends to the full range of human roles and responsibilities embedded within socio-technical infrastructures [4]. While it builds upon existing domains such as "usable security" and "human factors in security", human-centred cybersecurity moves beyond these to foreground the human not merely as a system component but as a constitutive force in shaping security dynamics. This shift entails recognising that users do not simply comply with or violate predefined security norms. Rather, they interpret, negotiate, and sometimes resist them [24]. For instance, design assumptions premised on individual device ownership may not hold in cultural contexts where device sharing is normative (see e.g., [10] for a case study on South Asian women's use of mobile phones). This calls for approaches enabling the design of systems that support shared use and reflect diverse values, expectations, and socio-cultural constraints, while also recognising the autonomy, dignity, and gendered sensitivities that may be implicated in such contexts.

Human-centred methodological approaches To date, cybersecurity research remains methodologically fragmented, with distinct disciplines offering disparate definitions, conceptual frameworks, and research priorities [7]. Such fragmentation has contributed to a limited development of shared understanding across the field. Human-centred cybersecurity addresses this fragmentation by adopting interpretive and context-sensitive methodologies that conceptualise security not as a purely technical objective, but as a situated social practice. This perspective foregrounds the everyday realities in which security is enacted and demands close attention to how individuals, organisations, states, and other actors perceive risk, navigate uncertainty, and engage with security practices within specific socio-technical contexts. To this end, human-centred cybersecurity often draws on qualitative and participatory approaches. Ethnographic studies, in-depth interviews, and participatory design workshops offer rich insights into the lived experience of security and reveal latent needs and values that are frequently overlooked by more traditional, technology-centred methods [8, 9]. These approaches, well established within the field of Human-Computer Interaction (HCI), also enable a more reciprocal relationship between researchers and participants, supporting the co-construction of knowledge and the collaborative definition of problems and priorities.

Technical approaches towards human-centred cybersecurity Previous (and much of the current) technical work aligned with human-centred cybersecurity has primarily focused on addressing human factors and enhancing usability. Efforts in this area include, among others, the development of more intuitive authentication interfaces, password management tools, and context-sensitive notification systems. Emerging research also explores formal modelling techniques to account for human variability (see e.g., [15]), seeking to bridge the gap between technical rigour and real-world complexity, offering a basis for more responsive and inclusive system design.

Further, to support the integration of human-centred principles within technical cybersecurity, Grobler et al., [4] propose a framework comprising three interrelated components, thus offering a structured lens through which to examine the complex interplay between users, technologies, and security practices. In this respect, *User components* focus on the ways individuals interact with and interpret security technologies. For example, in smart home environments, authentication mechanisms must be sensitive to the presence of multiple users and reflect varying roles, routines, and levels of access. *Usage components* encompass the technical and regulatory structures intended to safeguard users. For example, in IoT-based health monitoring, includes the implementation of clear and transparent privacy protections. And finally, *Usability components* addresses the accessibility and understandability of security features, aiming to reduce cognitive burden and promote sustained engagement with secure practices.

4.6.3 Challenges and Future Work

Despite the above and other recent advances, substantial challenges remain in advancing the human-centred cybersecurity agenda. These challenges are not solely technical in nature; they extend across conceptual, methodological, and sociocultural domains:

Challenge 1: Continued fragmentation of the field Realising the full potential of human-centred cybersecurity requires sustained collaboration across domains including computer science, human-computer interaction, psychology, sociology, law, and design. Interdisciplinary dialogue is essential to build shared vocabularies and methodological frameworks capable of addressing the complexity of socio-technical systems [7]. Without such integration, efforts to centre human experience in cybersecurity risk remain peripheral and insufficient.

Challenge 2: Narrow conceptualisation of the "user" Much of existing work implicitly privileges the end-user, neglecting other key actors involved in shaping cybersecurity practices and infrastructures [4]. Developers, system administrators, policy-makers, and security professionals, even bad actors also enact critical decisions that influence security outcomes. Human-centred cybersecurity must therefore broaden its analytical scope to account for these diverse stakeholders and the socio-technical contexts in which security work unfolds.

Challenge 3: Balancing security and usability/user experience Users are often faced with trade-offs between adhering to secure practices and maintaining efficiency in their everyday activities. Complex or overly rigid security protocols can lead to user disengagement, circumvention, or non-compliance, ultimately undermining the intended security outcomes [6]. Future research should therefore prioritise the development of adaptive systems that integrate security seamlessly into users' workflows and that respond dynamically to contextual cues and user intent.

Challenge 4: Inclusive security design Equity and inclusivity remain underdeveloped dimensions in most cybersecurity frameworks. Conventional models frequently assume a universal, idealised user, thereby marginalising those whose needs, practices, and constraints fall outside this narrow conception (see e.g., [10]). Human-centred cybersecurity therefore demands an intersectional lens-one that makes visible how identity, power, and access shape the lived experience of security, and that guides the design of systems which are not only technically robust, but also socially responsive, inclusive, and empowering.

Challenge 5: Accounting for emotional and psychological dimensions of cybersecurity Security incidents-or even ambiguous signals that suggest potential breaches [22] - can provoke anxiety, fear, and distrust, contributing to what has been termed 'cybernoia'. These effects are particularly acute in intimate settings such as the home, where the presence of IoT devices renders security breaches more opaque and unsettling. Future systems must therefore do more than simply prevent breaches or improve usability. They must also communicate clearly, support users in making sense of ambiguous or uncertain situations, and recognise the emotional labour involved in managing (in)security, underlining the importance of empathetic design, accessible feedback mechanisms, and robust support structures.

Challenge 6: Cybersecurity as a matter of public trust and democratic governance Security is not solely a technical function; it is also a societal value. Participatory approaches that involve users, communities, and civil society organisations in the design, deployment, and oversight of secure systems can enhance transparency, accountability, and legitimacy. Recasting cybersecurity as a collective, civic responsibility challenges the dominant emphasis on individual responsibility and opens up for more inclusive, resilient forms of governance; one that foregrounds shared accountability across the entire cybersecurity ecosystem.

Addressing the above challenges necessitates a critical re-examination of dominant paradigms and the development of interdisciplinary frameworks that centre on inclusivity, contextual sensitivity, and shared responsibility.

4.6.4 References

[1] Dourish, Paul and Grinter, Rebecca E and Delgado de la Flor, Jessica and Joseph, Melissa, "Security in the wild: user strategies for managing security as an everyday, practical problem," *Personal and ubiquitous computing*, vol. 8, no. 6, pp. 391–401, Nov. 2004, doi:10.1007/s00779-004-0308-5.

[2] Garfinkel, S. and Lipford, H.R., "Usable security: History, themes, and challenges," Morgan & Claypool Publishers, 2014. Available: https://books.google.com/books?id=HPS9BAAAQBAJ

[3] Alferaidi, Ali et al., "Challenges in human centric intelligent systems for wireless sensor networks: A state of art," *Transactions on Emerging Telecommunications Technologies*, Wiley, 2023, doi:10.1002/ett.4850.

[4] Grobler, M., Gaire, R., and Nepal, S., "User, usage and usability: Redefining human centric cyber security," *Frontiers in Big Data*, vol. 4, p. 583723, 2021, doi:10.3389/fdata.2021.583723.

[5] Kumaraguru, P. et al., "Protecting people from phishing: the design and evaluation of an embedded training email system," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2007, doi:10.1145/1240624.1240760.

[6] Zimmermann, V., and Renaud, K., "Moving from a 'human-as-problem' to a 'human-as-solution' cybersecurity mindset," *International Journal of Human-Computer Studies*, vol. 131, 2019, doi:10.1016/j.ijhcs.2019.05.005.

[7] Dunn Cavelty, M., "Cybersecurity Research Meets Science and Technology Studies," *Politics and Governance*, vol. 6, no. 2, pp. 22–30, 2018, doi:10.17645/pag.v6i2.1385.

[8] Dunphy, P. et al., “Understanding the experience-centeredness of privacy and security technologies,” in *Proceedings of the 2014 New Security Paradigms Workshop*, ACM, 2014, doi:10.1145/2683467.2683475.

[9] Renaud, K. and Flowerday, S., “Contemplating human-centred security & privacy research: Suggesting future directions,” *Journal of Information Security and Applications*, vol. 34, 2017, doi:10.1016/j.jisa.2017.05.006.

[10] Sambasivan, N. et al., “"Privacy is not for me, it’s for those rich women": Performative privacy practices on mobile phones by women in South Asia,” *SOUPS*, 2018. Available: https://www.usenix.org/conference/soups2018/presentation/sambasivan

[11] A. Kitkowska, M. Warner, Y. Shulman, E. Wästlund, and L. A. Martucci, “Enhancing privacy through the visual design of privacy notices: Exploring the interplay of curiosity, control and affect,” in *Proc. 16th Symp. Usable Privacy and Security (SOUPS)*, Zenodo, 2020, pp. 437–456. Available: https://doi.org/10.5281/ZENODO.3980478

[12] A. McCormac, T. Zwaans, K. Parsons, D. Calic, M. Butavicius, and M. Pattinson, “Individual differences and Information Security Awareness,” *Computers in Human Behavior*, vol. 69, pp. 151–156, Apr. 2017. Available: https://doi.org/10.1016/j.chb.2016.11.065

[13] R. W. Proctor and J. Chen, “The role of human factors/ergonomics in the science of security: Decision making and action selection in cyberspace,” *Human Factors*, vol. 57, no. 5, pp. 721–727, 2015. Available: https://doi.org/10.1177/0018720815585906

[14] I. S. Winkler and B. Dealy, “Information security technology? Don’t rely on it: a case study in social engineering,” in *Proc. 5th USENIX UNIX Security Symposium*, 1995, pp. 1–1. Available: https://www.usenix.org/publications/library/proceedings/security95/full_papers/winkler.pdf

[15] L. Viganò, “Formal methods for Socio-technical security: (formal and automated analysis of security ceremonies),” in *Lecture Notes in Computer Science*, Springer, 2022, pp. 3–14. Available: https://doi.org/10.1007/978-3-031-08143-9_1

[16] S. Garfinkel and H. R. Lipford, “Usable security: History, themes, and challenges,” Morgan & Claypool Publishers, 2014. Available: https://books.google.com/books?hl=en&lr=&id=HPS9BAAAQBAJ

[17] Y. Yao, J. R. Basdeo, S. Kaushik, and Y. Wang, “Defending my castle: A co-design study of privacy mechanisms for smart homes,” in *Proc. 2019 CHI Conf. Human Factors in Computing Systems*, ACM, 2019, Art. 255, pp. 1–12. Available: https://doi.org/10.1145/3290605.3300428

[18] S. Prange, C. George, and F. Alt, “Design considerations for usable authentication in smart homes,” in *Proc. Mensch und Computer*, ACM, 2021, pp. 311–324. Available: https://doi.org/10.1145/3473856.3473878

[19] V. Loscri, I. Symeonidis, M. Griesbacher, V. Deniau, D. Andreoletti, A. Chiumento, V. Dimitrova, I. Corradini, et al., “Interdisciplinary Security Aspects of Next-Generation Wireless Networks and Systems: BEiNG-WISE: State of Research and Future Research Steps,” Tech. Rep., 2025. Available: https://beingwise.eu/publications/deliverables/first-year-deliverable/

[20] D. Jampen, G. Gür, T. Sutter, and B. Tellenbach, “Don’t click: towards an effective anti-phishing training. A comparative literature review,” *Human-centric Computing and Information Sciences*, vol. 10, no. 1, Article 33, 2020. Available: https://doi.org/10.1186/s13673-020-00237-7

[21] K. D. Nguyen, H. Rosoff, and R. S. John, “Valuing information security from a phishing attack,” *Journal of Cybersecurity*, vol. 3, no. 3, pp. 159–171, 2017. Available: https://doi.org/10.1093/cybsec/tyx006

[22] A. Rostami, M. Vigren, S. Raza, and B. Brown, “Being hacked: Understanding victims’ experiences of IoT hacking,” in *Proc. 18th Symp. Usable Privacy and Security (SOUPS)*, 2022, pp. 613–631. Available: https://www.usenix.org/conference/soups2022/presentation/rostami

[23] B. Brown, M. Vigren, A. Rostami, and M. Glöss, “Why users hack: Conflicting interests and the political economy of software,” *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW2, Article 354, Nov. 2022. Available: https://doi.org/10.1145/3555774

[24] P. Sanches, V. Tsaknaki, A. Rostami, and B. Brown, “Under surveillance: Technology practices of those monitored by the state,” in *Proc. 2020 CHI Conf. Human Factors in Computing Systems*, ACM, 2020, pp. 1–13. Available: https://doi.org/10.1145/3313831.3376889

4.7 Security Challenges, Solutions, and Future Research Directions in IoMT

Authors: Gul Tahaoglu[1], Beste Ustubioglu[1], Guzin Ulutas[1]

[1] Department of Computer Engineering, Karadeniz Technical University, Türkiye

4.7.1 Introduction

The evolving role of networked health devices and platforms—collectively referred to as the *Internet of Medical Things* (IoMT)—has transformed healthcare delivery. From wearable, real-time monitors to hospital information systems that automate data management, IoMT promises greater efficiency, focused care, and reduced costs for providers and patients alike [1].

With these promises come a host of **security and privacy concerns**. IoMT ecosystems are highly distributed; each device and service can be a potential target for malicious actors seeking to compromise confidentiality, integrity, or availability of sensitive patient information. Encryption standards, identity and access management (IAM), and blockchain-based audit trails have been proposed to mitigate risk [2]. However, these measures are insufficient unless intrusions are detected promptly.

Consequently, intrusion detection systems (IDSs) form a critical layer of defense. Traditional signature-based IDSs struggle with emerging threats, while anomaly-based IDSs - particularly those employing machine-learning (ML) or deep-learning (DL) models—offer adaptive protection by learning what constitutes normal IoMT traffic and flagging deviations in real time [3]. These data-driven approaches demand realistic, domain-specific datasets for effective training and validation, which are challenging to obtain because of patient-privacy regulations.

This paper surveys state-of-the-art anomaly-based IDSs, complementary technologies such as blockchain, and open challenges that motivate future research in secure IoMT deployment.

4.7.2 State of the Art

Recent years have witnessed a surge of research activity committed to the development of more advanced security controls for IoMT environments, driven by both the growing sophistication of cyberattacks and the sensitivity of patient health data. Anomaly-based intrusion detection systems (IDS) have been particularly salient as a promising solution, offering a more adaptive defense layer compared to traditional signature-based ones. The following is an overview of prevailing trends and directions of state-of-the-art research.

Machine-Learning and Deep-Learning Approaches Supervised algorithms such as support-vector machines (SVMs) and random forests classify benign versus malicious traffic, while DL models—autoencoders, convolutional neural networks (CNNs), and recurrent architectures—capture complex dependencies in high-dimensional IoMT data. These models typically achieve higher detection precision and resilience against zero-day attacks.

Federated Learning and Privacy-Preserving Techniques Because IoMT data are geographically dispersed, *federated learning* (FL) allows edge devices to train a global anomaly detection model without centralising sensitive health records, thereby complying with HIPAA / GDPR while reducing data-transfer overhead [4].

Hybrid and Ensemble IDS Architectures Combining signature- and anomaly-based detection, often through ensemble learning, yields layered protection and improved false-positive or true-positive rates. Hybrid frameworks can dynamically adapt to novel attack vectors in evolving threat landscapes.

Blockchain Integration for Data Integrity Distributed ledgers provide tamper evident logs of device interactions and security events. When paired with anomaly-based IDS outputs, blockchain enables immutable forensic trails that strengthen compliance and incident response capabilities.

4.7.3 Challenges and Future Research Directions

Realistic Dataset Scarcity Strict privacy laws limit open access to clinical data, impeding the development and benchmarking of anomaly-detection models. Synthetic-data generation and privacy-preserving data-sharing frameworks (e.g. differential privacy) are promising avenues to bridge this gap.

Resource-Constrained Devices Many IoMT sensors possess limited processing power and battery life, constraining IDS complexity. Lightweight models or edge-/fog-computing offload strategies are essential to maintain detection accuracy without overburdening devices.

Adversarial Robustness ML-based IDSs are vulnerable to adversarial examples that manipulate input traffic to evade detection. Future work should explore robust training methods, continuous model updates, and defense-in-depth strategies to mitigate such attacks.

Blockchain Scalability and Interoperability Although blockchain enhances data integrity, public-ledger consensus protocols may introduce latency and storage overhead. Scalable Layer-2 solutions and cross-chain interoperability must mature before large-scale clinical adoption.

Explainability and Human-in-the-Loop Security Clinicians require transparent decision logic to trust AI-driven IDS alerts. Explainable-AI techniques and intuitive visual dashboards will facilitate effective human oversight and faster incident response.

4.7.4 Conclusion

Anomaly-based IDSs, blockchain audit trails, and privacy-preserving learning paradigms jointly constitute a compelling blueprint for secure IoMT ecosystems. Addressing dataset scarcity, resource constraints, adversarial robustness, and explainability will be pivotal for translating these advances into practical, clinically viable solutions.

4.7.5 References

[1] S. M. R. Islam, D. Kwak, M. H. Kabir, M. H. Hossain, and K.-S. Kwak, "The Internet of Things for Health Care: A Comprehensive Survey," *IEEE Access*, vol. 3, pp. 678–708, 2015. Available: https://doi.org/10.1109/ACCESS.2015.2437951

[2] X. Zheng, Y. Zhao, H. Li, R. Chen, and D. Zheng, "Blockchain-based verifiable privacy-preserving data classification protocol for medical data," *Computer Standards & Interfaces*, vol. 82, artcile, 103605, 2022. Available: https://doi.org/10.1016/j.csi.2021.103605

[3] P. Shojaei, E. Vlahu-Gjorgievska, and Y. W. Chow, “Security and privacy of technologies in health information systems: A systematic literature review,” *Computers*, vol. 13, no. 2, article 41, 2024. Available: https://doi.org/10.3390/computers13020041

[4] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, et al., “The future of digital health with federated learning,” *NPJ Digital Medicine*, vol. 3, no. 1, article 119, 2020. Available: https://doi.org/10.1038/s41746-020-00323-1

4.8 Wireless Network Security in Aviation: Threats, Challenges and Future Directions

Authors: Jawad Manzoor[1], Waqas Ahmed[1]

[1] University of Galway, Ireland

4.8.1 Introduction

Global air traffic has grown significantly in recent years due to rising demand, which leads to busier skies and more congested airspace. Additionally, the increasing use of drones in the commercial sector as well as for surveillance and military purposes, has added a new layer of complexity to air traffic management. Therefore, ensuring the security and efficiency of wireless communication networks used in aviation has become increasingly critical. Automatic Dependent Surveillance-Broadcast (ADS-B) is a next-generation aviation technology that uses GPS to provide accurate information about an aircraft's position and flight path. Unlike traditional radar systems, ADS-B enables aircraft to broadcast their location, speed, and other key details not only to air traffic controllers (ATC) but also to nearby planes. This real-time data sharing significantly enhances situational awareness and improves flight safety. It is mandatory for aircraft to have ADS-B in the majority of US and European controlled airspace. ADS-B has two main parts: ADS-B OUT, which automatically sends out information about an aircraft's location, speed, and altitude, and ADS-B IN, which lets aircraft receive similar information from others nearby. It mostly uses the radio frequencies of 1090 MHz and 978 MHz for communication. The International Civil Aviation Organization (ICAO) manages it globally, the Federal Aviation Administration (FAA) in the United States, and EUROCONTROL in Europe.

Despite its benefits, ADS-B has several security weaknesses due to its open and unencrypted nature. ADS-B messages are broadcast publicly without authentication or encryption. The system was designed with interoperability and widespread access as priorities rather than security. This open and unencrypted nature of ADS-B makes it vulnerable to various air-based and ground-based threats. Eavesdropping is one of the most common threats to ADS-B messages. Attackers can perform traffic analysis, data harvesting, and correlation attacks to breach confidentiality. The integrity of ADS-B communication can be compromised by modifying legitimate ADS-B messages. E.g., an attacker can modify the positional data sent by an aircraft so that it appears to be at false coordinates. Another major threat to ADS-B is the Denial of Service (DoS) because it makes ADS-B messages unavailable to authorized entities. An attacker can disrupt the communication between aircraft or ATCs by using various jamming techniques. ADS-B is also vulnerable to spoofing and message injection attacks due to a lack of authentication. E.g., an attacker can create a ghost aircraft that does not exist in reality but appears as an actual entity for ADS-B receivers. These attacks can confuse and mislead pilots and ATCs.

4.8.2 State of the Art

Aviation security has seen significant advancements, with a growing focus on addressing vulnerabilities in the ADS-B system. Recent developments in Machine Learning (ML) and Deep Learning (DL)-based anomaly detection methods have shown success in detecting malicious ADS-B messages. Supervised learning methods such as Support Vector Machines (SVM), Decision Trees, and k-nearest neighbors (KNN) are robust in classification tasks. In [1], a classification framework leveraging KNN, logistic regression, and naïve Bayes models is proposed. Using a dataset generated via the OpenSky API, the proposed techniques are evaluated against various attack types. The KNN model has been shown to outperform the others in terms of classification accuracy. DL models like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) have also been explored. These models are particularly good at understanding the temporal and spatial relationships of flight paths. LSTM has been used for spoofing attack detection by preprocessing ADS-B message sequences with a sliding window technique [3]. Adversarial learning techniques have also been used, with one study achieving 98% accuracy in detecting spoofed ADS-B data [7]. Additionally, CNNs have been employed for aircraft classification based on ADS-B signals. Graph Neural Networks (GNNs) are also effective in making anomaly detection more context-aware by treating air traffic as a network of interconnected elements.

Cryptographic methods like encryption, digital signatures, and authentication codes have also been studied for securing ADS-B communications and ensuring data integrity, authenticity, and confidentiality [11]. Lightweight encryption methods, such as Elliptic Curve Cryptography (ECC) and hybrid cryptographic techniques, are particularly well-suited for ADS-B systems, which are usually resource-constrained. These methods try to mitigate threats like message injection, spoofing, and tampering. One notable framework, ADS-Bsec [4], enhances security through a key management module that facilitates secure message authentication and integrity verification during flight transitions across air traffic control zones. Another approach [6] proposes a lightweight symmetric cryptography-based protocol, ensuring message integrity and confidentiality.

Some research works have investigated radio frequency fingerprinting techniques that exploit the distinct transmission characteristics of individual aircraft, such as frequency, signal strength, and timing variations [8]. These methods help in the identification of legitimate ADS-B transmitters and distinguish them from spoofed or malicious signals. Other methods rely on GPS data and multilateration techniques to identify inconsistencies in reported positions or times, and mitigate threats such as trajectory spoofing and ghost injection attacks. Kalman filter-based techniques predict and verify aircraft trajectories by analyzing both historical and real-time ADS-B data. They can detect anomalies such as sudden deviations or unrealistic flight paths by comparing observed positions with expected values [10].

4.8.3 Challenges and Future Work

While the innovations in ADS-B are promising, there are still many challenges in ensuring real-time processing and minimizing false positives. The non-availability of comprehensive datasets for ML and DL models' training is also a challenge. While real-world benign ADS-B data is readily available through projects like OpenSky, it is hard to collect sufficient labeled attack data. Keeping in view the current challenges facing ADS-B, the following are some future research directions.

- **Lightweight Privacy-Preserving Broadcasts:** Cryptographic methods can provide privacy and security to ADS-B communications, but they require changes to the system's message structure, which can be challenging to implement across global networks. Balancing privacy with transparency is required for effective air traffic management. To this end, techniques like pseudonymization and dynamic ICAO identifiers to protect sensitive information can be explored. Another promising approach is the lightweight encryption of certain aircraft data and sharing of decryption keys only with authorized parties. These methods aim to safeguard privacy without undermining the utility of ADS-B systems.

- **Distributed Anomaly Detection Networks** Instead of relying solely on isolated ground stations for anomaly detection, future research should focus on collaborative anomaly detection methods. Techniques such as federated learning and swarm intelligence can be utilized, where multiple nodes share insights without exposing raw data. This not only improves detection accuracy but also makes it harder for attackers to exploit vulnerabilities.

- **Blockchain for Trust Management** Blockchain technology offers more than just a way to log messages. It could be used for building trust across the aviation ecosystem. Smart contracts could automate data validation, manage access control, and resolve disputes when conflicting data streams arise. This decentralized approach ensures greater transparency and reliability.

- **Automated Incident Response** To stay ahead of cyber threats, the aviation industry needs automated countermeasures that can respond in real-time. Developing standardized incident response playbooks tailored to aviation environments could also help reduce delays in decision-making during cyber incidents.

- **Digital Twins for Security Testing** Digital twins that are virtual replicas of aviation environments can be invaluable for testing and evaluating potential cyber threats. These simulations allow researchers to experiment with attack scenarios and assess the effectiveness of defense mechanisms without risking real-world operations. It's a safe and cost-effective method to stay prepared for emerging threats.

- **Human Factors in Cybersecurity** While technology plays an important role in detecting and responding to cyber threats, human operators remain at the heart of aviation security. Understanding the human factors involved such as cognitive load, alert fatigue and interface design is essential. Better training programs and decision-support systems can empower ATC personnel and pilots to detect and respond to cyber threats more effectively.

4.8.4 References

[1] S. Khan, J. Thorn, A. Wahlgren, and A. Gurtov, "Intrusion detection in automatic dependent surveillance-broadcast (ADS-B) with machine learning," in *Proc. IEEE/AIAA 40th Digital Avionics Systems Conf. (DASC)*, 2021, pp. 1–10. Available: https://doi.org/10.1109/DASC52595.2021.9594431

[2] J. Wang, Y. Zou, and J. Ding, "ADS-B spoofing attack detection method based on LSTM," *EURASIP J. Wireless Commun. Networking*, vol. 2020, no. 1, 2020. Available: https://doi.org/10.1186/s13638-020-01756-8

[3] R. Karam, M. Salomon, and R. Couturier, "Supervised ADS-B anomaly detection using a false data generator," in *Proc. 2nd Int. Conf. Computer, Control and Robotics (ICCCR)*, 2022, pp. 218–223. Available: https://doi.org/10.1109/iccr54399.2022.9790149

[4] T. Kacem, A. B. Barreto, P. Costa, and D. Wijesekera, "A key management module for secure ADS-B," in *Proc. IEEE 25th Int. Conf. Intelligent Transportation Systems (ITSC)*, 2022, pp. 1784–1789. Available: https://doi.org/10.1109/ITSC55140.2022.9922200

[5] X. Chen, D. He, C. Peng, M. Luo, and X. Huang, “A secure and effective hierarchical identity-based signature scheme for ADS-B systems,” *IEEE Trans. Aerospace and Electronic Systems*, vol. 60, no. 4, pp. 5157–5168, 2024. Available: https://doi.org/10.1109/TAES.2024.3386148

[6] B. Sher, M. Ahmad, K. Mansoor, Y. A. Bangash, W. Iqbal, and S. Mussiraliyeva, “Lightweight secure authentication protocol for automatic dependent surveillance broadcast system,” *Cluster Computing*, 2024. Available: https://doi.org/10.1007/s10586-024-04566-5

[7] N. S. Joseph, C. Banerjee, E. Pasiliao, and T. Mukherjee, “FlightSense: A spoofer detection and aircraft identification system using raw ADS-B data,” in *Proc. IEEE Int. Conf. Big Data*, 2020, pp. 3885–3894. Available: https://doi.org/10.1109/BigData50022.2020.9377975

[8] G. Gurer, Y. Dalveren, A. Kara, and M. Derawi, “A radio frequency fingerprinting-based aircraft identification method using ADS-B transmissions,” *Aerospace*, vol. 11, no. 3, p. 235, 2024. Available: https://doi.org/10.3390/aerospace11030235

[9] A. Darabseh, S. Khandker, and C. Pöpper, “Securing the sky: Detecting aircraft location drifting through cross-checking receiver-based estimated and received ADS-B trajectories,” *Journal of Open Aviation Science*, vol. 1, no. 2, 2023. Available: https://doi.org/10.59490/joas.2023.7504

[10] M. Leonardi and G. Sirbu, “ADS-B crowd-sensor network and two-step Kalman filter for GNSS and ADS-B cyber-attack detection,” *Sensors*, vol. 21, no. 15, p. 4992, 2021. Available: https://doi.org/10.3390/s21154992

[11] A. Braeken, “Holistic air protection scheme of ADS-B communication,” *IEEE Access*, vol. 7, pp. 65251–65262, 2019. Available: https://doi.org/10.1109/ACCESS.2019.2917793

Chapter 5

Intelligence at the Edge: AI and Federated Learning

5.1 Introduction

Chapter Editors: Guzin Ulutas [1], Nicolas Sklavos [2], Mustafa A. Mustafa [3, 4]

[1] Department of Computer Engineering, Karadeniz Technical University, Türkiye
[2] Computer Engineering and Informatics Department (CEID), University of Patras, Greece
[3] The University of Manchester, UK
[4] COSIC, KU Leuven, Belgium

Federated learning (FL) is a distributed learning paradigm that allows participating devices to train a common model while keeping their data private. FL redefines the scalability of artificial intelligence (AI). In particular, ultra-low latency services envisaged by 6G and beyond networks require fast processing of high volumes of sensitive data, which brings AI capabilities directly to the edge to overcome the bandwidth limitations and privacy risks of centralized architectures. As a result, AI and FL are transforming wireless networks into adaptive and intelligent ecosystems.

This section discusses approaches to decentralized learning, adversarial defense, and privacy preservation that equip systems to detect threats, optimize resources, and deliver secure services dynamically. It starts by highlighting that, as 6G aims to integrate AI into next-generation wireless networks (NGWNs), this deep integration creates new vulnerabilities, tackling which would require integrated and robust defense mechanisms. It also examines the two-fold impact of agent-enabled large language models (LLMs) in wireless network-based cybersecurity. On one hand, LLMs can be used as tools by attackers to reduce the cost and increase the scalability of attacks; on the other hand, they can also be used as defensive tools, such as for automatic vulnerability discovery. The remaining part of the section focuses on FL, covering the potential and the risks to user privacy when deploying FL in NGWNs. It also reviews several secure and robust FL approaches by (i) highlighting the potential attack vectors and defense mechanisms, (ii) investigating ways to adapt FL-based intrusion detection systems to heterogeneous and resource-constrained environments, and (iii) highlighting the need to develop backdoor defenses that integrate client participation. The section also addresses vulnerabilities of FL by proposing a deception-based defense mechanism against poisoning attacks - a proactive defense strategy that aims to mislead attackers, consume their resources, and hinder their progress. Lastly, the section emphasizes that FL's multi-round training and communication cycle can lead to high energy consumption in IoT devices with limited battery life, reducing participation and increasing device drops; therefore, energy efficiency stands out as a critical research focus for the success of FL-based IDSs for IoT.

When evaluated together, a holistic picture emerges ranging from the wireless ecosystem's vision of embedding AI at every layer of the architecture with 6G to the potential of LLMs to automate the attack-defense cycle; from the privacy, heterogeneity, and energy constraints of FL-based intrusion detection systems approaches for zero-day detection in IoT, to FL's requirements for advanced reliability, cyber deception, and hyperparameter-independent defenses against poisoning and backdoor threats. The common finding is that an FL solution that keeps data local and is based on distributed intelligence is critical for privacy-compliant and scalable security; however, it also comes with extensive attack surfaces, variable attack rates, high energy consumption, and model integrity issues. Therefore, sustainable security in the future 6G-IoT environment seems to depend on the co-design of energy-efficient FL algorithms running at the edge, dynamic trust management involving the client, differential privacy, cyber deception mechanisms, and alignment-check layers that prevent the misuse of LLMs.

5.2 6G Network Specific Attacks Against AI Models

Authors: Sunder Ali Khowaja[1], Kapal Dev[2], Gurjot Singh Gaba[3]

[1] School of Computing, Dublin City University, Ireland

[2] Department of Computer Science, Munster Technological University, Ireland

[3] Department of Computer and Information Science (IDA), Linköping University, Sweden

5.2.1 Introduction

The introduction of a 6G communication network is poised to redefine wireless communications by not only addressing the problems associated with its predecessor fifth generation (5G) network, but also by integrating AI methods across multiple network layers. Unlike 5G, 6G does not use AI merely as an add-on, but embeds AI in the architecture seamlessly from edge computing nodes and radio access networks to centralized core elements. For instance, AI models can be deployed in distributed controllers such as RAN intelligent controllers (RIC) in OpenRAN (O-RAN) architecture that allows the network to perform security management, network slicing, and enable real-time decision making for resource allocation. The AI integration also facilitates adaptive user experience, dynamic optimization, and autonomous network management while significantly increasing system responsiveness and energy efficiency.

Although the integration of AI brings a lot of improvement in terms of performance, 6G's pervasive integration and reliance on AI introduces new vulnerabilities. The attacks concerning AI models in 6G target the very algorithms that underpin the intelligence systems by potentially compromising the security and reliability of the network. For example, poisoning attacks can target the neural receivers deployed at the edge by manipulating the training data. Such attacks can lead the AI models to misclassify the results that might lead to inefficient resource allocation or missed intrusion detections. The evasion attacks are responsible for adding subtle adversarial perturbations during the inference of AI model, thus, causing anomaly detection systems to overlook malicious traffic. Additionally, model inversion attacks are capable of recovering sensitive training data from AI model while the model extraction attacks allow adversaries to reconstruct proprietary AI models by querying the output. These type of attacks expose confidential network information, thus making the network vulnerable. Furthermore, Trojan attacks complicate the treat landscape concerning AI models through embedding hidden triggers during the training process. Such triggers are dormant until they are activated to disrupt critical network functions by inducing unauthorized behavior.

The distributed, virtualized, and open architecture of 6G networks characterized by multi-access edge computing, real-time edge processing, and loosely coupled components expand the attack surface considerably. With AI models being the integral part of 6G networks serving as the backbone for autonomous network operations, an attack could undermine the privacy of the whole network while degrading performance, disrupting services, and propagating the attack vulnerability throughout the network. Furthermore, AI model attacks are not always isolated incidents. The attacks can be systematic threats designed to exploit the inherent vulnerabilities in 6G architecture. In this paper, we highlight some state-of-the-art AI model-related attacks, examine their techniques, and discuss potential impacts within the 6G ecosystem. The paper also highlights the critical need for integrated and robust defense mechanisms to cope with the AI-model attacks.

5.2.2 State of the Art

With the progression of performance and capability of AI-based architectures, researchers have tried to integrate the functionalities to 6G communication systems in a seamless manner. For instance, researchers tried to leverage the distributed learning characteristics for 6G communication systems to improve the lifetime of the edge devices while reducing the latency of the systems in compliance to MEC architecture [1]. Some researchers leverage the split learning framework to not only improve the performance, but also to improve the security aspect of the 6G network [2]. Nevertheless, with the progression of performance, security vulnerabilities also increase from AI model perspective as the attackers can recreate the data through model inversion attacks or manipulate the data through model poisoning attacks, respectively. Some researchers try to combine the characteristics of generative AI to deal with model inversion and poisoning attacks in the recent literature [3], but it increases the latency which does not reflect the real-time characteristics. Below, we have summarized some state-of-the-art methods that understand and mitigate attacks specifically targeting AI models deployed in 6G networks.

- Poisoning attacks: Through this attack, the malicious actor adds subtle adversarial perturbations during inference leading AI systems to produce incorrect outputs. For instance, the study in [3] performs a model inversion attack to recreate the training data and then uses it to inject adversarial perturbations into the training process. The variation seems small, but with each communication round, the performance of the model decreases significantly. Another study in [4] demonstrated that poisoning attacks can bias neural network-based channel estimators, causing misallocation of resources in 6G edge nodes.

- Evasion attacks: Evasion attacks are quite similar to the poisoning attacks such that both these attacks involve adding adversarial perturbations. However, the poisoning attack targets training process while the evasion attack focuses on the inference process. The evasion attack leads the AI system to infer incorrect output. The study in [5, 6] highlighted the effect of evasion attack suggesting that even minor modifications can cause drastic misclassifications, which can compromise network security or crash the system for sensitive applications such as healthcare.

- Model extraction attacks: The attack focuses on the recreation of architecture and its parameters by exploiting the query interface of an AI model. The study in [7, 8] shows that proprietary AI models can be replicated effectively by iterative querying. Once the AI models can be recreated, it can be further used for targeted attacks, specifically in the cases of automated network management systems.

- Model inversion attacks: This type of attack lets the adversaries create sensitive training data by inverting the model gradients. The study in [8] highlights that the inversion techniques are capable of exposing private use data and confidential network parameters, especially when the model is deployed at distributed edge nodes.

- Trojan attack: The attack involves camouflaging hidden triggers during the training phase so that the malicious functionality can be activated with specific input behaviors that trigger the attack. The study in [5, 9] illustrated that the trojan attack on AI models in 6G networks can lead to unauthorized control of network slicing mechanism upon the trigger of specific behavior.

The above-mentioned attacks highlight the systemic vulnerabilities of AI model security and privacy state that can undermine the entire 6G network. The security challenges are

Table 5.1: Summary of different attack types and their characteristics and common point of attack in 6G MEC architecture.

Attack Type	Key Characteristics	Common Point of Attack in 6G MEC architecture
Poisoning Attack	- Manipulation of training Data - Can occur as model poisoning (compromising the training process) or data poisoning (malicious) samples.	- Targets distributed data aggregation nodes at the edge where IoT devices supply unverified data. - Exploits vulnerabilities in local training pipelines.
Evasion Attack	- Introduction of subtle adversarial perturbations during inference. - Causes misclassifications in real-time	- Attacks on MEC-based inference services where data transmitted from edge devices over less secure channels can be perturbed. - Exploits unreliable channel conditions.
Model Extraction Attack	- Reconstructs model parameters or internal architecture by querying model APIs. - Leaks Intellectual property.	- Edge servers open expose APIs for rapid inference; repeated querying can be exploited where access controls are less strict. - Vulnerable in Mobile Edge Computing environments.
Model Inversion Attack	- Infers sensitive training data from model outputs - Leads to privacy leakage of user and network data.	- MEC nodes, which aggregate data from diverse sources, may have insufficient sanitization. -Attackers exploit inference endpoints at the edge to recover confidential data.
Trojan Attack	- Embeds hidden triggers during training that remain dormant until activated. - Can cause unauthorized behavior upon activation.	- Edge training environments with limited monitoring are prone to injecting backdoors. - Once deployed, triggers in edge-based models can be remotely activated under specific conditions.

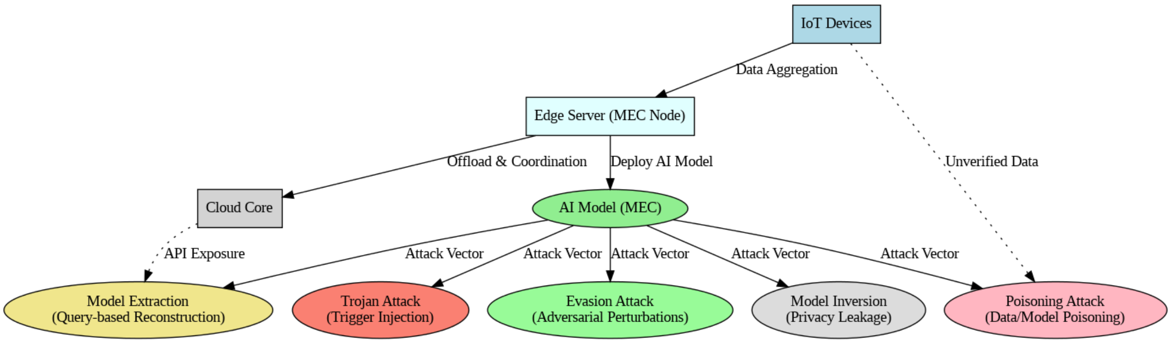


Figure 5.1: Key MEC Layers (IoT Devices, Edge Server, Cloud Core) and associated AI attacks.

further compounded by the virtualized and distributed architecture of 6G, where AI models are deployed across heterogeneous devices. The inherent interconnectivity and openness that is promised for the autonomous operation of 6G network enlarges the attack surface, thus making the robust security countermeasures indispensable. We summarize the key characteristics of the attack types in Table 5.1. We also provide the specific point of attack where these vulnerabilities commonly manifest in the distributed architecture of 6G networks. The aforementioned attacks are also shown in Figure 5.1.

5.2.3 Challenges and Future Work

It has been established well that the deployment of AI models is crucial to the 6G network. However, such a deployment poses security challenges concerning management and optimization of the communication system. For instance, given the distributed nature of MEC, edge nodes are vulnerable where the AI models are deployed. The key challenges here include how to avoid vulnerable data aggregation and poisoning, real-time evasion in resource constrained environment, exposed inference APIs and model extraction, privacy leakage via model inversion, and Trojan insertion in decentralized training.

We provide the following future work directions in a brief manner that could provide promising defense avenues against AI model attacks.

- End-to-end secure pipelines: Comprehensive pipelines need to be designed that integrate robust data filtering, real-time anomaly detection, and differential privacy measures to protect training data at the edge.
- Adaptive defense mechanisms: Dynamic model monitoring and continuous adversarial training need to be implemented in order to detect and respond to evasion attacks concerning resource-constrained MEC environments.
- Secure API framework: Secure API interfaces for AI model inference and update mechanisms need to be developed and standardized that incorporate encryption and rate limiting. Such measures would deter inversion and extraction attacks.
- Distributed Trojan detection: Design of collaborative frameworks that enable the nodes in the MEC architecture to detect anomalous behavior while sharing insights on model performance are needed. Such continuous monitoring would flag the behavior indicative of Trojan backdoors, thus ensuring consistent model integrity across the network.

We believe that the aforementioned challenges can be addressed by the future directions that are laid out. The possible research directions would not only ensure the security and privacy of AI model and users' data, but also ensure consistent model integrity across the 6G network.

5.2.4 Conclusion

The integration of AI within 6G networks offers unparalleled advantages in terms of efficiency, adaptability, and automation. However, this reliance on AI introduces critical security challenges that threaten network integrity and user privacy.

This study highlights key attack vectors, such as poisoning, evasion, model extraction, model inversion, and Trojan attacks, which exploit vulnerabilities in AI-based network functions. Given the distributed architecture of 6G and the increasing complexity of cyber threats, the traditional security measures are insufficient. To mitigate these risks, we propose building comprehensive security frameworks that incorporate end-to-end encryption, dynamic adversarial training, robust API security, and decentralized anomaly detection. By implementing these proactive strategies, future 6G networks can ensure the confidentiality, integrity, and availability of AI-driven operations.

The findings of this study underscore the need for continued research into AI model security to ensure a secure and resilient 6G ecosystem capable of withstanding evolving cyber threats.

5.2.5 References

[1] S. A. Khowaja, K. Dev, P. Khowaja, and P. Bellavista, "Toward energy-efficient distributed federated learning for 6G networks," *IEEE Wireless Communications*, vol. 28, no. 6, pp. 34–40, Dec. 2021. Available: https://doi.org/10.1109/MWC.012.2100153

[2] S. A. Khowaja, P. Khuwaja, K. Dev, K. Singh, L. Nkenyereye, and D. Kilper, "ZETA: ZEro-trust attack framework with split learning for autonomous vehicles in 6G networks," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, 2024, pp. 1–6. Available: https://doi.org/10.1109/WCNC57260.2024.10571158

[3] S. A. Khowaja, P. Khuwaja, K. Dev, and A. Antonopoulos, "SPIN: Simulated poisoning and inversion network for federated learning-based 6G vehicular networks," in *Proc.*

IEEE International Conference on Communications (ICC), 2023, pp. 6205–6210. Available: https://doi.org/10.1109/ICC45041.2023.10279339

[4] M. A. Ferrag, B. Kantarci, L. C. Cordeiro, M. Debbah, and K.-K. R. Choo, "Poisoning attacks in federated edge learning for digital twin 6G-enabled IoTs: An anticipatory study," in *Proc. IEEE ICC Workshops*, 2023, pp. 1253–1258. Available: https://doi.org/10.1109/ICCWorkshops57953.2023.10283797

[5] B. D. Son, N. T. Hoa, T. V. Chien, W. Khalid, M. A. Ferrag, W. Choi, and M. Debbah, "Adversarial attacks and defenses in 6G network-assisted IoT systems," *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 19168–19187, Jun. 2024. Available: https://doi.org/10.1109/JIOT.2024.3373808

[6] K. Zerhouni, G. S. Gaba, M. Hedabou, T. Maksymyuk, A. Gurtov, and E. M. Amhoud, "GAN-based evasion attack in filtered multicarrier waveforms systems," *IEEE Trans. Machine Learning in Communications and Networking*, vol. 2, pp. 210–220, 2024. Available: https://doi.org/10.1109/TMLCN.2024.3361834

[7] M. A. Ferrag, O. Friha, B. Kantarci, N. Tihanyi, L. Cordeiro, M. Debbah, D. Hamouda, M. Al-Hawawreh, and K.-K. R. Choo, "Edge learning for 6G-enabled Internet of Things: A comprehensive survey of vulnerabilities, datasets, and defenses," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2654–2713, 2023. Available: https://doi.org/10.1109/COMST.2023.3317242

[8] S. A. Khowaja, P. Khuwaja, K. Dev, K. Singh, X. Li, N. Bartzoudis, and C. R. Comsa, "Block encryption LAyer (BELA): Zero-trust defense against model inversion attacks for federated learning in 5G/6G systems," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 807–819, 2025. Available: https://doi.org/10.1109/OJCOMS.2025.3526768

[9] S. Bhunia, M. S. Hsiao, M. Banga, and S. Narasimhan, "Hardware Trojan attacks: Threat analysis and countermeasures," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1229–1247, Aug. 2014. Available: https://doi.org/10.1109/JPROC.2014.2334493

5.3 LLMs and Wireless Attacks: Threats and Opportunities

Authors: Piotr Przymus[1], Andreas Happe[2], Jürgen Cito[2]

[1] Faculty of Mathematics and Computer Science, Nicolaus Copernicus University in Toruń, Poland
[2] TU Wien, Vienna University of Technology, Austria

5.3.1 Introduction

This White Paper explores the dual-edged nature of large language models (LLMs), esp. agentic LLMs, in wireless-based cyber attacks. LLMs can not only enhance but fully automate the cyber attack lifecycle—spanning reconnaissance, initial access, privilege escalation, lateral movement, maintaining access, and impact on target. While LLMs can be used offensively, they also offer defensive benefits. Organizations can harness LLMs to detect vulnerabilities, simulate realistic attack scenarios, and proactively strengthen their security posture.

To make it more concrete, we analyze the Nearest Neighbor attack [3], where an advanced persistent threat actor (APT) used Wi-Fi networks to bypass multi-factor authentication (MFA), showing how LLMs could be used to scale and refine similar attacks.

Case-study: nearest neighbor attack. Discovered in 2024, this attack targeted Company A, whose VPN access was protected by MFA. To bypass it, the attackers first compromised a nearby Company B within wireless range. After breaching Company B's network, they accessed a dual-homed device connected to both wired and wireless networks. This allowed them to infiltrate Company A's wireless network—both internal and guest Wi-Fi—which, unlike VPN, lacked MFA protection, granting internal access. Further investigation revealed they also routed through another nearby Company C to mask their movements. This attack required significant time, expertise, and infrastructure to penetrate multiple networks before reaching the final target. While LLMs were not involved, they could be used to automate key stages, reducing both effort and cost for attackers by automating time-consuming tedious tasks.

5.3.2 State of the Art

LLMs for offensive security. Attack methodologies outline the phases of an attack. This study follows the Mandiant Attacker Life Cycle (shown in Figure 5.2), emphasizing iterative control loops—attackers do not rely on a single exploit, but repeatedly adapt within the target network to achieve their goal. During the initial *Reconnaissance* phase, attackers collect relevant information about their target, e.g., attackers can use passive OSINT methods to gather target data, typically combining tools like Shodan, SpiderFoot, and leaked credential databases. Then in the *Initial Access* phase, access to the target network is gained, e.g., by using phishing, attacking vulnerable external services, or by using leaked credentials [1]. Later, the attacker employs *Privilege Escalation* to gain privileges, e.g., user or service accounts, and use *Lateral Movement* techniques to traverse the target network or access additional connected internal networks. Attackers often use command-and-control (C2) frameworks, e.g., CobaltStrike or Sliver, to *Maintain Persistence* in their target network. Finally, during the *Impact upon Target* phase, attackers profit from their attack through, e.g., industrial espionage, denial-of-service, or ransomware attacks.

Using a case study analyzed through this framework, we identify areas where LLMs can already enhance attacker efficiency:

- **Reconnaissance.** PassGAN [5] uses deep learning to expand leaked password lists.

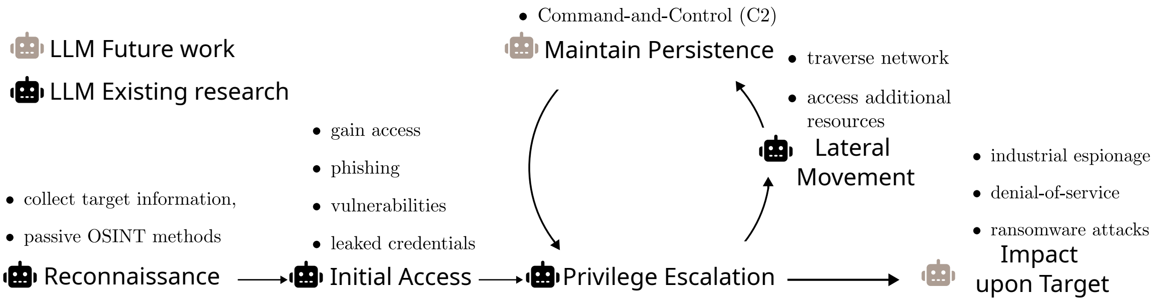


Figure 5.2: Mandiant attacker life cycle, with phases explanation and information about potential LLM usage.

Recent research indicates that LLMs already contain sufficient background information and are able perform open-source intelligence (OSINT) work [4].

- **Initial Access.** Zhou et al. [6] use LLMs to automatically find and exploit vulnerabilities, while Heiding at el. investigate the feasibility of using LLMs for automated spear phishing [7], highlighting the decreased cost and increased scalability of LLM-based attacks.
- **Lateral Movement** & **Privilege Escalation.** Happe et al. used LLMs in this matter in both Linux systems [9] and Microsoft Active Directory enterprise networks [8]. Singer et al. have shown that LLMs are capable of traversing through multiple networks to achieve their goal [10].

LLMs for defensive security Given enough trust and safe-guards, an LLM could be employed to perform system and network hardening to proactively prevent attackers from abusing these vulnerable configurations. While the usage of LLMs decreases the cost of attacks for adversaries, it also allows defenders to proactively test their own networks. LLMs can also be utilized to automate configuration hardening and fix software vulnerabilities. [11].

Note Basic cyber-security hygiene will prevent most of traditional and LLM based attacks. Employing 802.1x together with device-identification as well as strong network segmentation would have broken this example attack chain.

5.3.3 Challenges and Future Work

LLMs are transforming cybersecurity on both the offensive and defensive fronts. While they enhance defensive capabilities, they also lower the barrier for attackers, making it essential to continuously research safeguards and mitigation strategies to stay ahead.

LLMs for offensive security LLMs can streamline “tedious reconnaissance” [12] by automating OSINT data collection with tools like Shodan and SpiderFoot, reducing manual effort through function-calling capabilities. Attackers use C2 frameworks like Sliver and CobaltStrike to maintain persistence, where LLMs can streamline onboarding and automate tasks. For financial gain or espionage, LLMs can efficiently process stolen data, with locally deployed models bypassing perimeter detection. In ransomware operations, LLM-driven chatbots could handle victim negotiations. Fully automated cyberattacks through reasoning models may become a reality.

LLMs for defensive security LLMs can be leveraged for network and behavioral analysis, typically via endpoint detection and response (EDR) tools, enabling earlier attack detection. With sufficient safeguards, LLMs could also automate system and network hardening to prevent exploitation. While LLMs lower attack costs for adversaries, they also help defenders by making penetration testing more accessible, reducing the cost of vulnerability discovery.

Future work The growing risk of LLM-driven attacks underscores the need for robust safeguards. Preventing locally-run LLMs from executing unintended malicious actions will be critical, requiring advancements in model alignment and controlled access to sensitive functions. Beyond individual safeguards, end-to-end defense strategies must evolve in response to automated threats. LLM-driven detection, response, and mitigation techniques will need to keep pace with increasingly sophisticated attack automation, ensuring that security measures remain resilient in this rapidly shifting landscape.

5.3.4 References

[1] CERT NZ, "How ransomware happens and how to stop it," Available: https://www.cert.govt.nz/information-and-advice/guides/how-ransomware-happens-and-how-to-stop-it/, Accessed: 2025-03-13.

[2] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," Technical Report, 2010. Available: https://www.ciosummits.com/media/solution_spotlight/LM_Cyber_Kill_Chain_White_paper_2011.pdf

[3] S. Adair, "The Nearest Neighbor Attack: How A Russian APT Weaponized Nearby Wi-Fi Networks for Covert Access," Volexity Blog, Nov. 2024. [Online]. Available: https://www.volexity.com/blog/2024/11/22/the-nearest-neighbor-attack-how-a-russian-apt-weaponized-nearby-wi-fi-networks-for-covert-access/, Accessed: 2025-03-13.

[4] A. Happe and J. Cito, "Getting pwn'd by AI: Penetration testing with large language models," in *Proc. 31st ACM Joint European Software Engineering Conf. and Symp. Foundations of Software Engineering*, ACM, Nov. 2023. Available: https://doi.org/10.1145/3611643.3613083

[5] B. Hitaj, P. Gasti, G. Ateniese, and F. Perez-Cruz, "PassGAN: A deep learning approach for password guessing," *arXiv [cs.CR]*, Sep. 2017. Available: https://doi.org/10.1007/978-3-030-21568-2_11

[6] Z. Zhou, Y. Yang, S. Wu, Y. Huang, B. Chen, and X. Peng, "Magneto: A step-wise approach to exploit vulnerabilities in dependent libraries via LLM-empowered directed fuzzing," in *Proc. 39th IEEE/ACM Int. Conf. Automated Software Engineering*, ACM, Oct. 2024, pp. 1633–1644. Available: https://doi.org/10.1145/3691620.3695531

[7] F. Heiding, S. Lermen, A. Kao, B. Schneier, and A. Vishwanath, "Evaluating large language models' capability to launch fully automated spear phishing campaigns: Validated on human subjects," *arXiv preprint arXiv:2412.00586*, Nov. 2024. Available: https://arxiv.org/abs/2412.00586

[8] A. Happe and J. Cito, "Can LLMs hack enterprise networks? Autonomous assumed breach penetration-testing Active Directory networks," *arXiv preprint arXiv:2502.04227*, Feb. 2025. Available: https://arxiv.org/abs/2502.04227

[9] A. Happe, A. Kaplan, and J. Cito, “LLMs as hackers: Autonomous Linux privilege escalation attacks,” *arXiv preprint arXiv:2310.11409*, Oct. 2023. Available: https://arxiv.org/abs/2310.11409

[10] B. Singer, K. Lucas, L. Adiga, M. Jain, L. Bauer, and V. Sekar, “On the feasibility of using LLMs to execute multistage network attacks,” *arXiv e-prints*, Jan. 2025. Available: https://ui.adsabs.harvard.edu/abs/2025arXiv250116466S/abstract

[11] Z. Ye, T. H. M. Le, and M. A. Babar, “LLMSecConfig: An LLM-based approach for fixing software container misconfigurations,” in *Proc. IEEE Working Conference on Mining Software Repositories*, Feb. 2025, pp. 629–641. Available: https://doi.org/10.1109/MSR66628.2025.00099

[12] A. Happe and J. Cito, “Understanding hackers’ work: An empirical study of offensive security practitioners,” in *Proc. 31st ACM Joint European Software Engineering Conf. and Symp. Foundations of Software Engineering*, ACM, Nov. 2023. Available: https://doi.org/10.1145/3611643.3613900

5.4 Federated Learning and Privacy in IoT Wireless Networks

Authors: Nesibe Yalçın[1]

[1] Department of Computer Engineering, Erciyes University, Türkiye

5.4.1 Introduction

Recent improvements in wireless communication have accelerated the broad use of the internet of things (IoT). In an IoT network, a large number of interconnected electronic devices and sensors operate in the background to collect user and environmental data. However, these devices are vulnerable to numerous cyber attacks where attackers can intercept and analyze sensitive/private data [1]. Moreover, device/data owners may be reluctant to share their data with a centralized entity because of security concerns such as data privacy and confidentiality [2]. Next-generation wireless networks (NGWNs), with their ultra-dense deployment, heterogeneous infrastructure, extremely complex and dynamic topology, further amplify these vulnerabilities, expand the attack surface, and expose IoT devices to more sophisticated threats.

Federated learning (FL) is a promising solution due to its inherent ability to preserve data and user privacy for IoT. In the FL approach, data entities/clients can collaborate to learn a global model jointly without sacrificing data privacy [3]. By enabling decentralized model training (allowing data to be processed locally on devices) and handling large amounts of data in real time, the risk of data breaches can be minimized and compliance with privacy regulations can be achieved simultaneously. In this paper, the potential of FL, the deployment of FL in the context of IoT networks, and the preservation of privacy based on FL are discussed.

5.4.2 State of the Art

Machine learning (ML) has been increasingly applied to cope with the dynamic nature of IoT [4] and meet the heterogeneous requirements of NGWNs [5], [6]. In the ML approach, data must be transmitted to a centralized entity for preprocessing and model training. This can significantly increase communication overloads and lead to privacy issues, especially since it requires sharing confidential data. FL, a disruptive distributed ML framework (as shown in Figure 5.3) first introduced by Google in 2016, focuses on the centralization of models rather than raw data, thus reducing the cost of communication and providing a solution to ensure data privacy (without disclosing their data to others) [7], [8]. With its communication efficiency and privacy-preserving features, FL has become a natural choice for NGWNs.

Several recent works employed FL methods to overcome or partially solve data privacy issues in various NGWNs such as sixth generation (6G) networks. Zarandi and Tabassum proposed a federated double deep Q-learning approach to enhance the learning speed of IoT/edge devices and minimize their privacy concerns [9]. On the other hand, FL methods continue to encounter security and privacy challenges [2]. For example, a central server may be vulnerable to tampering or hacking by personnel, thus model parameters/updates may be leaked or corrupted. Zhu et al. [2] have proposed a blockchain-based FL framework for IoT supply chain management. This framework also aims to ensure the data's resistance to tampering and to maintain data security in the FL.

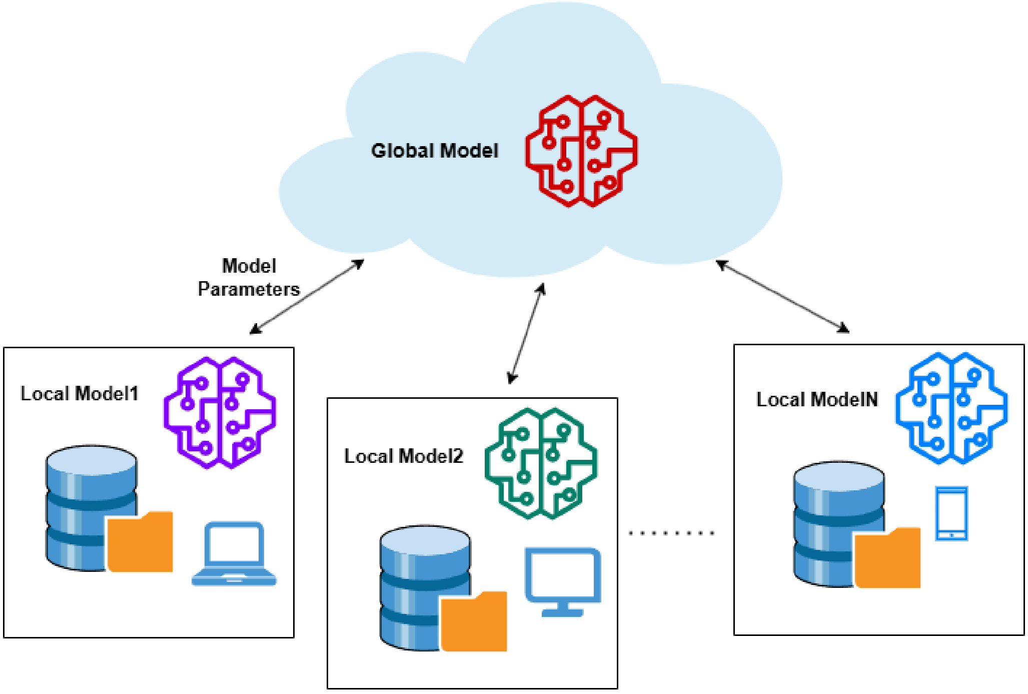


Figure 5.3: Illustration of FL.

5.4.3 Challenges and Future Work

FL is an emerging approach to make communication more efficient, to reduce the computational load, and to address privacy preservation challenges for IoT devices (agents) and wireless networks. This offers significant advantages, particularly for the privacy protection of personal data [10]. Nevertheless, its performance depends on conditions of the wireless channels and it can still be vulnerable to various cyber threats and privacy issues owing to the involvement of many end-users [5].

FL can be further enhanced by integrating with blockchain and encryption to provide stronger protection and significantly improve data privacy, particularly considering privacy constraints. In addition, the authenticity of model updates can be verified through a robust mechanism. Differential privacy mechanisms can be applied to local updates, thus protecting individual contributions to the model and further enhancing privacy. With these advancements, this study concludes with the motivation and insights to leverage FL methods for data security and privacy in wireless networks.

5.4.4 References

[1] N. Mishra and S. Pandya, “Internet of things applications, security challenges, attacks, intrusion detection, and future visions: A systematic review,” *IEEE Access*, vol. 9, pp. 59353–59377, 2021. Available: https://ieeexplore.ieee.org/abstract/document/9405669/

[2] L. Zhu, S. Hu, X. Zhu, C. Meng, and M. Huang, “Enhancing the security and privacy in the IoT supply chain using blockchain and federated learning with Trusted Execution Environment,” *Mathematics*, vol. 11, no. 17, p. 3759, 2023. Available: https://doi.org/10.3390/math11173759

[3] O. Aouedi, K. Piamrat, G. Muller, and K. Singh, "Federated semisupervised learning for attack detection in industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 286–295, 2022. Available: https://ieeexplore.ieee.org/abstract/document/9729433/

[4] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1996–2018, 2014. Available: https://ieeexplore.ieee.org/abstract/document/6805162/

[5] D. Shome, O. Waqar, and W. U. Khan, "Federated learning and next generation wireless communications: A survey on bidirectional relationship," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 7, e4458, Jul. 2022. Available: https://doi.org/10.1002/ett.4458

[6] I. Ahmad, S. Shahabuddin, H. Malik, E. Harjula, T. Leppänen, L. Loven, A. Anttonen, A. H. Sodhro, et al., "Machine learning meets communication networks: Current trends and future challenges," *IEEE Access*, vol. 8, pp. 223418–223460, 2020. Available: https://ieeexplore.ieee.org/abstract/document/9274307/

[7] J. C. Jiang, B. Kantarci, S. Oktug, and T. Soyata, "Federated learning in smart city sensing: Challenges and opportunities," *Sensors*, vol. 20, no. 21, p. 6230, 2020. Available: https://doi.org/10.3390/s20216230

[8] Y. Canbay and Y. Büyüknacar, "Federe Öğrenme ve Veri Mahremiyeti," in *Yapay Zekâ ve Büyük Veri Çalışmaları, Siber Güvenlik ve Mahremiyet* (Eds: Ş. Sağıroğlu, M. U. Demirezen), Nobel Yayınevi, 2021.

[9] S. Zarandi and H. Tabassum, "Federated double deep Q-learning for joint delay and energy minimization in IoT networks," in *Proc. IEEE International Conference on Communications Workshops (ICC Workshops)*, Jun. 2021, pp. 1–6. Available: https://doi.org/10.1109/iccworkshops50388.2021.9473821

[10] L. Zang, X. Zhang, and B. Guo, "Federated deep reinforcement learning for online task offloading and resource allocation in WPC-MEC networks," *IEEE Access*, vol. 10, pp. 9856–9867, 2022. Available: https://doi.org/10.1109/ACCESS.2022.3144415

5.5 Secure Federated Learning: A Model Analysis Approach for Mitigating Malicious Participants

Authors: Eda Sena Erdol[1], Hakan Erdol[2], Beste Ustubioglu[1], Guzin Ulutas[1], Iraklis Symeonidis[3]

[1] Karadeniz Technical University, Türkiye
[2] University of Bristol, UK
[3] RISE, Research Institutes of Sweden, Sweden

5.5.1 Introduction

Federated learning (FL) [1] represents an innovative approach that enables the training of machine learning models across distributed devices while preserving user privacy. This paradigm maintains data privacy by processing sensitive information without centralised data collection at the server, simultaneously reducing communication overhead. The approach has gained particular importance in domains that require strict data protection, such as healthcare, finance, and mobile applications, and has become increasingly significant with the implementation of regulatory frameworks such as the general data protection regulation (GDPR).

FL systems can be categorised based on both network architecture and data partitioning aspects. Centralised frameworks employ a single server that coordinates model aggregation and updates, offering faster convergence but introducing single-point-of-failure risks. Decentralised architectures enable peer-to-peer communication without a central server, eliminating single-point vulnerabilities but increasing communication overhead. Most research prioritises centralised frameworks for security analysis because defending against threats in decentralised architectures would require trusting potential attackers—an impractical assumption when any participant could be malicious.

Most state-of-the-art studies consider a cetralised FL framework for problem formulation and defense against malicious threats. One of the main reasons is that each user in the FL network can be a potentially malicious user. Therefore, defending in a decentralised structure requires trusting in a potential source of the attack. Moreover, to defend against such attacks in a decentralised architecture, all users should contribute to the defence including the malicious ones, which can be considered as an unrealistic assumption.

Figure 5.4 illustrates the taxonomy of FL attack vectors, data distribution scenarios and defence mechanism among state-of-the-art studies.

Due to its distributed architecture, the FL framework is prone to potential vulnerabilities to various adversarial attacks, including data poisoning and model poisoning, which are classified based on the adversary's objectives and implementation methodologies. These security threats can significantly compromise the accuracy and reliability of the model, potentially compromising the entire FL infrastructure. The following sections present these threats and related defensive measures with particular emphasis on centralised FL frameworks.

5.5.2 State of the Art

Attack Vectors in FL

Research into FL security has uncovered a complex landscape of vulnerabilities and protective measures, with particular relevance to wireless deployments. Integrity attacks targeting the training process reveal several patterns based on their implementation approaches and underlying objectives.

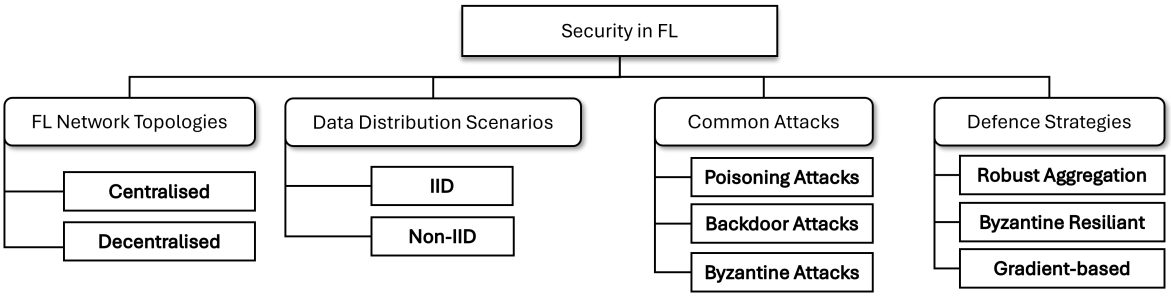


Figure 5.4: FL Security Taxonomy.

Model poisoning attacks Model gradient-level manipulation has proven significantly harmful to FL networks by directly altering model weights [2]. In model poisoning, malicious participants intentionally modify their local weights before sharing them, allowing attackers to influence model with subtle adjustments. Zhou et al. [3] refined these methods by developing covert approaches that target specific network components, substantially complicating detection through traditional comparison methods.

Data poisoning attacks While comparatively less efficient, training data poisoning remains a threat to wireless FL environments. Researchers categorise these into preserved and flipped label approaches. The former introduces carefully crafted forged data without changing classifications, often employing noise or synthetic data generation techniques. The latter directly transforms classifications through fixed or adaptive relabeling strategies. Recent advances include distance-aware techniques that measure feature space relationships between categories to optimise effectiveness while reducing detection probability.

Studies comparing attack methodologies commonly demonstrate that direct parameter manipulation achieves higher success rates with significantly fewer compromised participants than data poisoning approaches, with model poisoning requiring approximately 5% of the participant infiltration needed for a comparable impact of data poisoning - a critical efficiency consideration in resource-constrained wireless environments [4].

Defense Mechanisms

Recent protective countermeasures generally follow three design strategies: contribution analysis, resilient aggregation techniques, and verification protocols.

Gradient-based contribution analysis This approach employs mathematical distance metrics to identify suspicious contributions. Techniques such as cosine similarity [5] and influence functions [4] measure the deviation of user updates from expected patterns. These methods are particularly effective against model poisoning attacks as they directly analyse gradient patterns, enabling early detection of manipulated model weights before they can significantly influence the global model.

Byzantine resilient methods These techniques modify traditional weight averaging through approaches such as Statistical filtering to remove outliers. Byzantine-resilient aggregation [6] have demonstrated effectiveness against targeted attacks while maintaining model performance on legitimate tasks. These methods demonstrate superior performance against targeted model poisoning attacks, though they require more computational resources than other defences.

Robust aggregation methods These techniques protect against both model and data poisoning by modifying the traditional averaging process to minimise the impact of malicious updates. Krum [7] selects updates with minimal Euclidean distance to others, effectively eliminating outliers. Trimmed Mean removes extreme values for each parameter before averaging, while coordinate-wise median aggregation replaces means with medians, offering Byzantine resistance with reasonable convergence guarantees. While effective against both model and data poisoning attacks, these methods show particular strength against data poisoning attacks by minimising the influence of outliers that typically result from poisoned data, all while maintaining reasonable convergence properties.

In wireless FL environments, naturally heterogeneous data distributions create additional detection challenges, as benign model variations often resemble malicious modifications. This fundamental conflict between privacy preservation and security requires innovative frameworks that address both concerns without requiring access to sensitive local information or imposing excessive computational demands on resource-limited devices.

5.5.3 Challenges and Future Work

Implementing robust security in wireless FL frameworks faces several critical challenges that require innovative research approaches. The computational demands of current defence mechanisms often exceed the capabilities of resource-constrained wireless devices, creating a need for lightweight security protocols that maintain effectiveness while reducing computation requirements.

Resource constraints The computational demands of current defense mechanisms often exceed the capabilities of resource-constrained wireless devices. This situation makes lightweight security protocols that maintain effectiveness while reducing processing requirements vital. Future research should focus on developing efficient anomaly detection algorithms specifically designed for low-power devices and investigating hardware-accelerated security mechanisms compatible with edge computing platforms.

Non-IID data challenges Non-uniform data distribution further complicates detection mechanisms, as legitimate variations in local models often resemble malicious manipulations. Therefore, non-IID data distributions prove to be a promising topic that requires developing more sophisticated anomaly detection techniques to distinguish between natural model divergence and adversarial manipulation in these heterogeneous environments.

Addressing these challenges requires interdisciplinary collaboration to develop secure and privacy-preserving federated learning systems for next-generation wireless intelligent applications.

5.5.4 References

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *International Conference on Artificial Intelligence and Statistics*, PMLR, vol. 54, pp. 1273–1282, 2016. Available: https://proceedings.mlr.press/v54/mcmahan17a?ref=https://githubhelp.com

[2] D. Yin, Y. Chen, K. Ramchandran, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," in *International Conference on Machine Learning*, PMLR, vol. 80, pp. 5636–5645, 2018. Available: https://proceedings.mlr.press/v80/yin18a

[3] X. Zhou, M. Xu, Y. Wu, and N. Zheng, "Deep model poisoning attack on federated learning," *Future Internet*, vol. 13, no. 3, p. 73, 2021. Available: https://doi.org/10.3390/fi13030073

[4] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, et al., "Advances and open problems in federated learning," *Foundations and Trends in Machine Learning*, vol. 14, no. 1-2, pp. 1–210, 2021. Available: http://dx.doi.org/10.1561/2200000083

[5] K. Pillutla, S. M. Kakade, and Z. Harchaoui, "Robust aggregation for federated learning," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1142–1154, 2022. Available: https://ieeexplore.ieee.org/abstract/document/9721118/

[6] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020. Available: https://ieeexplore.ieee.org/abstract/document/9084352/

[7] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," in *Neural Information Processing Systems*, vol. 30, pp. 119–129, 2017. Available: https://proceedings.neurips.cc/paper/2017/hash/f4b9ec30ad9f68f89b29639786cb62ef-Abstract.html

5.6 Adversarial Robust Defence Technique for Zero-day Attacks against Federated IoT Devices

Authors: Chibueze Peace Obioma[1], Youcheng Sun[1,2], Mustafa A. Mustafa[1,3]

[1] The University of Manchester, UK
[2] MBZUAI, UAE
[3] COSIC, KU Leuven, Belgium

5.6.1 Introduction

The deployment of billions of interconnected IoT devices facilitates enhanced monitoring, real-time analytics, and unprecedented operational efficiencies. However, this expanding connectivity landscape has simultaneously broadened the attack surface, making IoT networks highly susceptible to new threats, particularly zero-day attacks [1]. Zero-day attacks exploit previously unknown vulnerabilities, rendering traditional signature-based intrusion detection systems (IDS) inadequate due to their reliance on known threat signatures. Therefore, innovative detection techniques are urgently needed.

Federated learning (FL) has emerged as a promising solution for IoT applications, enabling collaborative model training across distributed IoT devices while preserving user privacy. This makes FL particularly suited for IDS in highly distributed IoT networks. To robustly defend against adversarial threats in federated IoT environments, various defense techniques (e.g., robust aggregation, anomaly detection, cryptographic protocols) have been proposed. Nonetheless, these solutions are often not effective in highly heterogeneous IoT environments.

Motivated by this critical gap, we examine how recent advancements in adversarial machine learning and cryptographic techniques can be leveraged to create robust FL-based IDS solutions. We critically analyze existing federated IDS approaches, zero-day detection methods, and adversarial resilience strategies, evaluating their applicability to IoT systems characterized by diverse device capabilities, limited computational resources, and strict privacy regulations.

5.6.2 State of the Art

Botnets and malware represent significant threats to IoT networks due to their ability to exploit vulnerabilities en masse. Diverse research [1, 3, 5, 6, 8] have extensively explored FL-based approaches for their detection. We provide a classification of state-of-the-art papers proposing FL-based IDS for IoT in Table 5.2.

Metwaly and Elhenawy [4] proposed FL-based solutions that incorporate privacy protection while defending against botnet. Rawat and Kumar [11] explored blockchain-enabled FL frameworks specifically designed for malware detection, demonstrating blockchain's potential for ensuring transparent, secure aggregation of federated models, thereby preventing malicious tampering or poisoning attacks during training phases. He et al. [10] demonstrated blockchain-powered FL with conditional generative adversarial networks (GANs) to adddress data scarcity and imbalance issues. Although blockchain-enhanced aggregation protocols have shown promise for secure and transparent aggregation, they incur high computational overhead and require significant adaptations.

The work in [7] used differential privacy (DP) and advanced federated aggregation methods to address privacy and heterogeneity challenges. DP-based mechanisms and secure multi-party computation have been explored extensively to bolster privacy protections. DP mecha-

Table 5.2: State-of-the-art papers proposing FL-based IDS for IoT.

Attack Type	IoT Environment	Data Handling	Papers
Zero-day, Botnet, General IDS	General IoT	Non-IID, IID	[1], [4]
General IDS, Malware	General IoT, Industrial IoT	Non-IID	[7]
Botnet	General IoT	Non-IID, IID	[3]
General IDS, Malware	General IoT, UAV-based IoT	Non-IID, IID	[10]
General IDS	General IoT	Non-IID	[8]
General IDS	General IoT, 5G Networks	Non-IID	[5]
General IDS	General IoT	Non-IID	[7]

nisms though inherently introduce noise, potentially reducing detection accuracy—especially against subtle, highly adaptive zero-day threats—highlighting the persistent challenge of balancing privacy preservation and security performance in federated IoT environments. Although methods such as DP, clustering-based aggregation, and data augmentation via GANs have been proposed to address heterogeneity, they are often computationally intensive, reducing their viability in IoT environments.

FL has also been explored as an adaptive mechanism to counter zero day attacks. The work by [3] emphasized FL's potential to identify anomalous patterns indicative of zero-day vulnerabilities by collaboratively analyzing heterogeneous traffic data across diverse IoT devices. Research result by [2] extensively demonstrated federated architectures' potential for proactive threat detection by leveraging aggregated intelligence from diverse IoT devices. For zero-day attack scenarios, effective detection fundamentally relies on accurate anomaly identification against normal traffic baselines, which becomes extremely challenging under non-IID conditions. The failure of these techniques to effectively mitigate zero-day exploits points to the fact that more research is needed.

5.6.3 Challenges and Future Direction

Below, we list some of the remaining open challenges in FL-based IDS for IoT.

Vulnerabilities in federated aggregation mechanisms One of the foremost challenges identified in federated IoT security contexts revolves around the inherent vulnerabilities within federated aggregation mechanisms. Adversarial participants may exploit aggregation protocols through malicious model updates or gradient poisoning attacks, significantly degrading the integrity of global detection models. Additionally, studies have identified the ineffectiveness of standard aggregation mechanisms against collusion-based attacks, where multiple adversaries synchronize their poisoned updates to bypass simple anomaly-detection filters. Thus, balancing security, computational efficiency, and scalability within federated aggregation mechanisms remains an open challenge.

Ensuring privacy and confidentiality under adversarial conditions Privacy preservation and confidentiality form foundational attributes of federated learning frameworks. However, adversaries increasingly leverage sophisticated inference attacks targeting federated models to infer sensitive information from aggregated model parameters. This problem becomes more acute under adversarial conditions associated with zero-day attacks, as anomalous behavior detection frequently requires more extensive visibility into traffic patterns and device states, inherently conflicting with stringent privacy constraints.

Data heterogeneity and non-IID distribution across IoT devices Another major challenge is the prevalent data heterogeneity and Non-IID nature of IoT datasets distributed across edge devices. IoT devices typically generate highly heterogeneous data distributions owing to their varying functionalities, sensor modalities, geographic locations, and user-specific usage patterns. Such heterogeneity profoundly impacts the federated training process, exacerbating model convergence difficulties and resulting in uneven detection performance across devices. For zero-day attack scenarios, effective detection fundamentally relies on accurate anomaly identification against normal traffic baselines, which becomes extremely challenging under non-IID conditions.

Resource constraints and limited computational capabilities The resource constrained nature of IoT devices introduces a fundamental barrier for implementing sophisticated adversarial defense techniques. IoT devices typically possess limited computational power, memory, storage, and battery life, rendering conventional deep learning models and computationally heavy defensive algorithms unsuitable for direct deployment. Popoola et al. [3] specifically highlighted that memory-efficient algorithms like LSTM-autoencoders, although promising, must further balance computational efficiency and accuracy, particularly when defending against unpredictable zero-day threats. Consequently, developing lightweight yet robust federated learning-based adversarial defenses capable of real-time, resource-efficient zero-day detection remains a challenge.

Future research directions A critical direction for future research involves developing more sophisticated federated aggregation methods capable of effectively thwarting adversarial threats, particularly those involving malicious model updates and parameter poisoning. Future methods must integrate advanced anomaly detection, blockchain transparency, and robust reputation-based scoring mechanisms to detect and neutralize threats proactively and accurately. Future research should also explicitly integrate adaptive context-aware mechanisms and meta-learning strategies to handle Non-IID data distributions effectively. Techniques that combine federated learning with meta-learning paradigms (e.g., few-shot learning) promise significant potential for rapidly adapting detection models to unseen zero-day attacks by leveraging limited anomaly samples across heterogeneous IoT devices. These methods can dynamically adjust models, improve generalization capabilities, and offer more effective protection against unknown and evolving adversarial threats.

5.6.4 References

[1] S. I. Popoola, R. Ande, B. Adebisi, G. Gui, M. Hammoudeh, and O. Jogunola, "Federated deep learning for zero-day botnet attack detection in IoT-edge devices," *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 3930–3944, 2021. Available: https://ieeexplore.ieee.org/abstract/document/9499122/

[2] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022. Available: https://ieeexplore.ieee.org/abstract/document/9751703/

[3] S. I. Popoola, B. Adebisi, M. Hammoudeh, H. Gacanin, and G. Gui, "Stacked recurrent neural network for botnet detection in smart homes," *Computers & Electrical Engineering*, vol. 92, article 107039, 2021. Available: http://dx.doi.org/10.1016/j.compeleceng.2021.107039

[4] A. Metwaly and I. Elhenawy, “Protecting IoT devices from BotNet threats: A federated machine learning solution,” *Sustainable Machine Intelligence Journal*, vol. 1, no. 1, pp. 15–25, 2023. Available: http://dx.doi.org/10.61185/smij.2023.22105

[5] L. Lavaur, B. Costé, M.-O. Pahl, Y. Busnel, and F. Autrel, “Federated learning as enabler for collaborative security between not fully-trusting distributed parties,” *Artificial Intelligence and Cybersecurity*, pp. 65–80, 2022. Available: https://imt-atlantique.hal.science/hal-03831515/

[6] E. Fedorchenko, E. Novikova, and A. Shulepov, “Comparative review of the intrusion detection systems based on federated learning: Advantages and open challenges,” *Algorithms*, vol. 15, no. 7, article 247, 2022. Available: http://dx.doi.org/10.3390/a15070247

[7] P. Ruzafa-Alcazar, P. Fernández-Saura, E. Mármol-Campos, A. González-Vidal, J. L. Hernández-Ramos, J. Bernal-Bernabe, and A. F. Skarmeta, “Intrusion detection based on privacy-preserving federated learning for the industrial IoT,” *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1145–1154, 2021. Available: https://ieeexplore.ieee.org/abstract/document/9609643/

[8] R. Yang, H. He, Y. Xu, B. Xin, Y. Wang, Y. Qu, and W. Zhang, “Efficient intrusion detection toward IoT networks using cloud–edge collaboration,” *Computer Networks*, vol. 228, article 109724, 2023. Available: http://dx.doi.org/10.1016/j.comnet.2023.109724

[9] S. I. Popoola, B. Adebisi, M. Hammoudeh, G. Gui, and H. Gacanin, “Hybrid deep learning for botnet attack detection in the Internet-of-Things networks,” *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4944–4956, 2020. Available: https://ieeexplore.ieee.org/abstract/document/9241019/

[10] X. He, Q. Chen, L. Tang, W. Wang, and T. Liu, “CGAN-based collaborative intrusion detection for UAV networks: A blockchain-empowered distributed federated learning approach,” *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 120–132, 2022. Available: https://ieeexplore.ieee.org/abstract/document/9863068/

[11] P. Rawat and P. Kumar, “Blockchain-based federated deep learning framework for malware attacks detection in IoT devices,” in *Proc. 14th Int. Conf. Computing Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2023, pp. 1–10. Available: http://dx.doi.org/10.1109/icccnt56998.2023.10306828

5.7 Enhancing Federated Learning Robustness Through Client Integration

Authors: Fatima Z. Abacha[1], Sin G. Teo[2], Lucas C. Cordeiro[1], Mustafa A. Mustafa[1,3]

[1] The University of Manchester, UK
[2] Institute for Infocomm Research, A*STAR, Singapore
[3] COSIC, KU Leuven, Belgium

5.7.1 Introduction

Federated learning (FL) enables multiple participants to collaboratively train a global model while preserving data confidentiality. A server coordinates training by broadcasting an initial model, which clients refine locally before sending updates for aggregation. This iterative process continues until convergence [1].

However, studies have shown that FL is vulnerable to security breaches caused by malicious participants. Adversarial clients may train local models on corrupted data or alter model parameters, compromising the global model. These attacks fall into two categories: untargeted, which degrade overall accuracy, and targeted, which cause misclassification of specific inputs. A specific form of targeted attacks is the *backdoor attack*, where an adversary manipulates its local model to induce specific labels when a trigger is present. Since backdoor attacks do not affect clean inputs, they are difficult to detect and mitigate.

FL also faces the well-known non-Identical and independently distributed (Non-IID) data heterogeneity challenge. Differences in data distribution, quality, and quantity among participants hinder model convergence to a global minimum. Moreover, data heterogeneity introduces fairness issues, as the global model often favors overrepresented groups, discouraging client participation in FL. Beyond impacting model performance and fairness, data heterogeneity also increases FL's vulnerability to security threats. When client data distributions vary significantly, it becomes harder to detect malicious behavior, making FL more susceptible to adversarial attacks.

5.7.2 State of the Art

Defenses against backdoor attacks in FL are mainly classified into two: (i) *pre-aggregation* defenses that focus on identifying malicious clients and prohibiting their participation in the global model aggregation, and (ii) *in/post-aggregation* methods, which mitigate the effect of backdoors in the aggregated global model.

Pre-aggregation defenses Pre-aggregation defenses adopt two approaches to distinguish between benign and malicious models: *robust aggregation* and *detection and isolation*.

Robust aggregation Robust aggregation approach is designed to maintain the integrity and performance of the global model despite the presence of adversarial participants. Blanchard et al. [2] introduced Krum and Multi-Krum, non-linear aggregation strategies that utilize Euclidean distances to select local models. They identify the local models most similar to the majority to form the basis of the global model. Another defense FoolsGold [3] operates on the assumption that malicious local models will exhibit similarity to each other, whereas benign clients will present more diverse model updates. Other defenses like FLTrust [4] utilize

Table 5.3: Comparison of Backdoor Attack Defense Methods in FL.

Defense	Strategy	Side	Limitation
Krum/Multi-Krum [2]	Euclidean Distance	Server	Vulnerable if malicious majority
FoolsGold [3]	Cosine Similarity	Server	Ineffective against diverse clients
FLTrust [4]	Auxiliary dataset	Server	Requires auxiliary dataset
FLDetector [5]	Cauchy mean theorem	Server	Partial client participation
FedRecover [6]	Use historical updates	Server	High storage & computational cost
VAE-based Detection [7]	Use reconstruction loss	Server	Complex training
FedCVAE [8]	Use reconstruction loss	Server	Complex training
Snowball [9]	Clustering and VAE	Server	Sensitivity to clustering accuracy
Gradient Pruning [10]	Pruning specific neurons	Server	Reduced benign task accuracy
Differential Privacy [11]	Add noise to reduce backdoor	Server	Security and performance trade-off
FLAME [12]	Clustering, clipping, noising	Server	Sensitivity to hyperparameters
FLIP [13]	Adversarial training	Client	Degradation of global accuracy
Crowd-Guard [14]	Client Feedback	Both	Requirement of TEE

an auxiliary dataset to train a reference model at the server side, which serves as a benchmark to evaluate the integrity of local models submitted by clients.

Detection and isolation Detection and isolation approach employs techniques to identify anomalous local models with the aim of isolating malicious contributions. The server in FLDetector [5] predicts model updates using the Cauchy mean value theorem. Clients whose updates deviate from these preditions receive a higher suspicion score and are classified as malicious. Similarly, FedRecover [6] maintains a copy of historical updates to produce a clean global model in the event of attacks. Li et al. [7] used a variational auto-encoder (VAE) to learn the pattern of the parameters of benign models trained in a centralized fashion. The trained VAE serves as a malicious client detector as the VAE outputted larger reconstruction losses for the unlearned patterns that are likely to be malicious. FedCVAE [8] reduces client updates to low-dimensional surrogate vectors and computes their geometric median. A CVAE is then used to compute reconstruction errors for the processed surrogate vectors, and local models with reconstruction errors above the mean are deemed adversarial and eliminated. A recent defense, Snowball [9], incorporates K-means clustering with VAE to identify benign clients.

In/post-aggregation defenses In/post-aggregation defenses typically commence during the aggregation of local models into a single global model. The server employs techniques such as gradient clipping to reduce the impact of the backdoor embedded in the poisoned models or gradient pruning, which deletes backdoor neurons after the aggregation phase. Wu et al. [10] proposed a gradient pruning method that identifies and erases neurons that are only triggered by backdoors. Naseri et al. [11] utilized differential privacy (DP) to effectively inhibit the effects of backdoors. Some studies have also proposed hybrid approaches that combines both pre- and post-aggregation strategies. In [12], Nguyen et al. developed FLAME, a three-part backdoor detector comprising clustering, clipping, and noising components that effectively defends against backdoor attacks in FL.

Client-side defenses Recent work [13, 14] proposed client-side adversarial training and utilizing client feedback to enhance the robustness of FL in the presence of adversaries, resulting in significant reduction in backdoor attack accuracy. Table 5.3 provides a comparative summary of the backdoor defenses.

5.7.3 Challenges and Future Work

Detecting backdoors in FL scenarios where participants possess highly non-IID data is not trivial. We highlight key challenges and directions for future research.

Client integration to reinforce server side defenses Current defenses mainly focus on mitigating attacks from the server side. Only a few approaches (FLIP [13], CrowdGuard [14]) integrate clients to enable defending the FL landscape. As these works show promising results, exploring more defense techniques that leverage client participation can be a valuable direction for future work.

Fixed threshold for adversaries Robust aggregation methods (as Krum [2]) are inherently designed to withstand a fixed percentage of malicious clients. As such, their effectiveness degrades significantly when the number of adversaries exceeds their designed tolerance. This necessitate the design of defenses that can tolerate flexible ratio of malicious candidates. Future work should focus on designing defenses that can flexibly handle varying proportion of adversaries.

Reliance on data assumptions and auxiliary data Some defenses depend on strong assumptions regarding data distribution or require access to an auxiliary dataset, limiting their practicality. For example, FoolsGold [3] performs well in non-IID settings but struggles under IID, while FLTrust [4] relies on the server possessing a trusted validation dataset. Real-world data is highly heterogeneous, and such assumptions may not hold. Future research should focus on designing methods that generalize across different data distributions and eliminate dependency on validation datasets. A promising direction is using high-quality synthetic data generated from foundation models to homogenize the data distribution of clients.

Expensive computational requirements Defenses that depend on historical updates and VAEs impose significant computational and communication overhead. This limits the deployment capability of these defenses in edge devices which usually operate in resource constrained environments.

Effects of hyperparameters The robustness of many existing defenses depend on specific choice of hyperparameters, such as the learning rate or backdoor target of the learning task. For example, incorporating DP without careful consideration can significantly deteriorate the global model accuracy. Similarly, FLAME [12] tends to misclassify benign models under highly non-IID scenarios and is sensitive to hyperparameters such as low learning rates. Recent approaches like Snowball [9], while resistant to attacks, results in reduced accuracy in cross-silo settings, and exhibit sensitivity to the choice of backdoor target. Thus, there is a critical need to develop hyperparameter-agnostic defense mechanisms to enhance the robustness of FL in the presence of adversaries.

5.7.4 References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," *International Conference on Artificial Intelligence and Statistics (AISTATS)*, PMLR, vol. 54, pp. 1273–1282, 2017. Available: https://proceedings.mlr.press/v54/mcmahan17a

[2] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," *NeurIPS*, vol. 30, pp. 119–129, 2017. Available: https://proceedings.neurips.cc/paper/2017/hash/f4b9ec30ad9f68f89b29639786cb62ef-Abstract.html

[3] C. Fung, C. J. Yoon, and I. Beschastnikh, "Mitigating sybils in federated learning poisoning," arXiv preprint arXiv:1808.04866, Aug. 2018. Available: http://arxiv.org/abs/1808.04866

[4] X. Cao, M. Fang, J. Liu, and N. Z. Gong, "FLTrust: Byzantine-robust federated learning via trust bootstrapping," arXiv preprint arXiv:2012.13995, Dec. 2020. Available: http://arxiv.org/abs/2012.13995

[5] Z. Zhang, X. Cao, J. Jia, and N. Z. Gong, "FLDetector: Defending federated learning against model poisoning attacks via detecting malicious clients," *Proc. 28th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, pp. 2545–2555, Aug. 2022. Available: http://dx.doi.org/10.1145/3534678.3539231

[6] X. Cao, J. Jia, Z. Zhang, and N. Z. Gong, "FedRecover: Recovering from poisoning attacks in federated learning using historical information," in *Proc. IEEE Symposium on Security and Privacy (SP)*, pp. 1366–1383, May 2023. Available: http://dx.doi.org/10.1109/sp46215.2023.10179336

[7] S. Li, Y. Cheng, W. Wang, Y. Liu, and T. Chen, "Learning to detect malicious clients for robust federated learning," arXiv preprint arXiv:2002.00211, Feb. 2020. Available: http://arxiv.org/abs/2002.00211

[8] Z. Gu and Y. Yang, "Detecting malicious model updates from federated learning on conditional variational autoencoder," in *Proc. IEEE Int. Parallel and Distributed Processing Symposium (IPDPS)*, pp. 671–680, May 2021. Available: http://dx.doi.org/10.1109/ipdps49936.2021.00075

[9] Z. Qin, F. Chen, C. Zhi, X. Yan, and S. Deng, "Resisting backdoor attacks in federated learning via bidirectional elections and individual perspective," *Proc. AAAI Conf. Artificial Intelligence*, vol. 38, no. 13, pp. 14677–14685, Mar. 2024. Available: http://dx.doi.org/10.1609/aaai.v38i13.29385

[10] C. Wu, X. Yang, S. Zhu, and P. Mitra, "Toward cleansing backdoored neural networks in federated learning," in *Proc. IEEE Int. Conf. Distributed Computing Systems (ICDCS)*, pp. 820–830, Jul. 2022. Available: http://dx.doi.org/10.1109/icdcs54860.2022.00084

[11] M. Naseri, J. Hayes, and E. De Cristofaro, "Local and central differential privacy for robustness and privacy in federated learning," arXiv preprint arXiv:2009.03561, Sep. 2020. Available: http://arxiv.org/abs/2009.03561

[12] T. D. Nguyen, P. Rieger, H. Chen, H. Yalame, H. Möllering, H. Fereidooni, S. Marchal, M. Miettinen, A. Mirhoseini, S. Zeitouni, F. Koushanfar, A. Sadeghi, and T. Schneider, "FLAME: Taming backdoors in federated learning," in *Proc. USENIX Security Symposium*, pp. 1415–1432, 2022. Available: https://www.usenix.org/conference/usenixsecurity22/presentation/nguyen

[13] K. Zhang, G. Tao, Q. Xu, S. Cheng, S. An, Y. Liu, S. Feng, G. Shen, P. Y. Chen, S. Ma, and X. Zhang, “FLIP: A provable defense framework for backdoor mitigation in federated learning,” arXiv preprint arXiv:2210.12873, Oct. 2022. Available: http://arxiv.org/abs/2210.12873

[14] P. Rieger, T. Krauß, M. Miettinen, A. Dmitrienko, and A. R. Sadeghi, “CrowdGuard: Federated backdoor detection in federated learning,” arXiv preprint arXiv:2210.07714, Oct. 2022. Available: http://arxiv.org/abs/2210.07714

5.8 Deception-Based Mitigation of Poisoning Attacks in Federated Learning

Authors: Grace Colette Tessa Masse[1], Abderrahim Benslimane[1]

[1] Centre d'Enseignement et de Recherche en Informatique (CERI), Université d'Avignon, France

5.8.1 Introduction

Artificial intelligence (AI) has become a cornerstone of modern technological advancement, driven by the proliferation of cloud-based applications and the ubiquitous presence of connected devices such as smartphones and tablets. These devices continuously generate vast amounts of data, fueling machine learning (ML) model training. However, traditional ML paradigms rely on centralized data aggregation, raising critical privacy and security concerns due to the risks of data leakage, surveillance, and single points of failure.

Federated learning (FL) has emerged as a decentralized paradigm that enables collaborative model training while preserving user data privacy. By retaining data on local devices and sharing only model updates, FL mitigates privacy issues linked to centralization. Nevertheless, this distributed nature introduces new security challenges, particularly in the form of poisoning attacks, where adversarial clients attempt to corrupt the global model either through manipulated training data (*data poisoning*) or by crafting harmful local model updates (*model poisoning*).

To defend against these threats, existing approaches can be categorized into:

1. *Model analysis*: Detects anomalous behavior in local models using statistical or learning-based techniques [1, 2].

2. *Byzantine-robust aggregation*: Adjusts the aggregation strategy (e.g., clipping, robust mean, or coordinate-wise median) to reduce the impact of outliers or malicious gradients [3, 4].

3. *Verification-based defense*: Verifies the integrity of updates via trusted execution environments (TEEs) or challenge-response tests [5, 6].

Although effective to some extent, these solutions are mostly reactive and aim to neutralize attacks without imposing significant cost on adversaries. This motivates the use of *cyber deception*, a proactive defense strategy adapted from cybersecurity, which aims to mislead attackers, consume their resources, and hinder their progress.

In this work, we propose a secure and adaptive FL framework that incorporates cyber deception alongside traditional defenses to enhance robustness against poisoning attacks. The framework (illustrated in Figure 5.5) includes:

- A **trust management strategy** based on the Dirichlet distribution to evaluate client behavior over time. Each client's trust score is dynamically updated according to the consistency and reliability of its model updates. The distribution's parameters are adjusted as new evidence accumulates, allowing the system to identify and down-weight untrustworthy participants.

- A **privacy-preserving and robust aggregation module** that integrates differential privacy (to limit information leakage during aggregation) and Byzantine-resilient techniques (to suppress outliers or malicious updates), both executed server-side to safeguard the global model integrity.

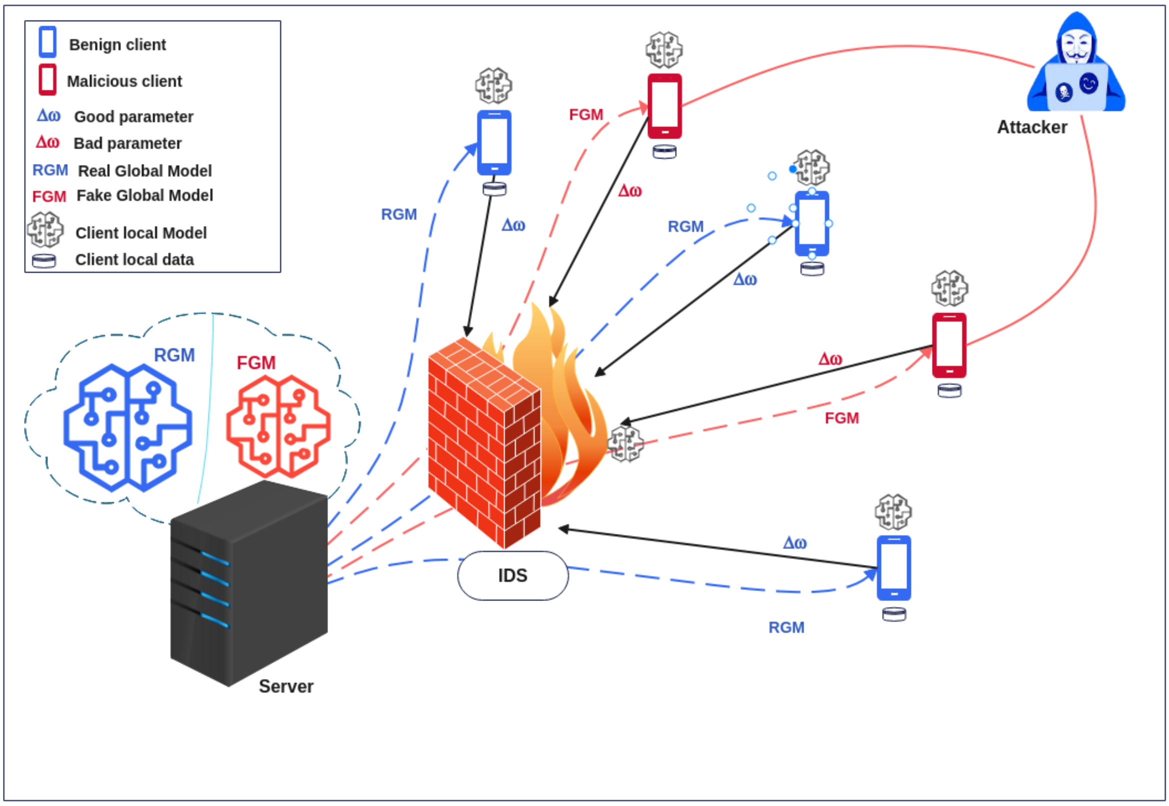


Figure 5.5: Overview of the proposed secure and deceptive FL framework.

- A novel **global decoy model (GDM)** that actively deceives attackers by simulating a realistic but isolated FL environment. Malicious clients identified or suspected via the trust mechanism are redirected to interact exclusively with the GDM.

The GDM is trained using both genuine and adversarial updates, but remains isolated from the real global model (RGM). It sends plausible but strategically manipulated gradients back to attackers to mislead their optimization efforts. By simulating model responses and preserving training plausibility, the GDM avoids detection while wasting attacker resources. Mechanisms are also in place to ensure benign clients are never exposed to the GDM, thereby maintaining service integrity.

By synergizing trust evaluation, robust aggregation, and adversarial deception, our framework offers a multi-layered defense that not only protects against poisoning but also strategically deters and drains malicious actors.

5.8.2 State of the Art

Existing defenses against poisoning in FL typically fall into three categories: model analysis, Byzantine-resilient aggregation, and verification-based methods.

Model analysis Model analysis approaches seek to detect poisoned models using statistical distance metrics or anomaly detection. For example, updates that deviate significantly from the expected gradient distribution may be flagged as suspicious [1].

Byzantine-resilient aggregation Byzantine-resilient aggregation techniques – such as coordinate-wise median, Krum, and norm bounding—aim to suppress the impact of outliers or adversarial gradients. Differential privacy is often combined with these methods to protect against information leakage, although it may reduce model utility [4].

Verification-based defenses Verification-based defenses leverage secure hardware (e.g., Intel SGX) or adversarial challenge-response schemes to validate update authenticity before aggregation [6].

However, these methods are generally passive, focusing on defense without retaliation. Deception-based strategies go further by actively engaging and misleading attackers. For instance, feeding adversaries with misleading model feedback or isolating them into fake training loops can drastically reduce their attack efficacy.

Our proposed framework builds on this insight by combining these defenses with an adaptive decoy mechanism that redirects and misguides malicious clients, turning their efforts against them while preserving model quality.

5.8.3 Challenges and Future Work

While our framework presents a promising direction for defending FL against poisoning, several challenges remain open:

- **Adaptive adversaries**: Attackers may evolve strategies to detect decoy environments. Future research could explore adversarial training to simulate and preempt such behaviors.
- **Decoy realism**: Ensuring the GDM is indistinguishable from the real model is critical. Further work is needed to refine gradient shaping and interaction timing to preserve the illusion.
- **Benign client safety**: A key concern is ensuring that no benign client mistakenly interacts with the GDM. Strict segregation protocols and dynamic whitelisting mechanisms must be maintained.
- **Scalability and efficiency**: Training and maintaining the GDM introduces computational overhead. Exploring lightweight deception strategies or resource-aware GDM variants could enhance deployability.
- **Beyond poisoning**: Although this work focuses on poisoning attacks, extending deception to defend against inference or model inversion attacks could further harden FL systems.
- **Game-theoretic integration**: Modeling the interaction between defenders and attackers as a dynamic game could help predict and counter adversarial adaptations, paving the way for more strategic and robust FL security mechanisms.

By addressing these challenges, future research can advance the development of intelligent, deceptive, and cost-efficient defenses for FL ecosystems operating in adversarial environments.

5.8.4 References

[1] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, and H. V. Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020. Available: https://ieeexplore.ieee.org/document/9069945

[2] A. Yazdinejad, A. Dehghantanha, H. Karimipour, G. Srivastava, and R. M. Parizi, “A robust privacy-preserving federated learning model against model poisoning attacks,” *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 6693–6708, 2024. Available: https://ieeexplore.ieee.org/document/10574838

[3] Z. Ma, J. Ma, Y. Miao, Y. Li, and R. H. Deng, “ShieldFL: Mitigating model poisoning attacks in privacy-preserving federated learning,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1639–1654, 2022. Available: https://ieeexplore.ieee.org/document/9762272

[4] X. Liu, H. Li, G. Xu, Z. Chen, X. Huang, and R. Lu, “Privacy-enhanced federated learning against poisoning adversaries,” *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4574–4588, 2021. Available: https://ieeexplore.ieee.org/document/9524709

[5] K. Li, J. Zheng, X. Yuan, W. Ni, O. B. Akan, and H. V. Poor, “Data-agnostic model poisoning against federated learning: A graph autoencoder approach,” *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 3465–3480, 2024. Available: https://ieeexplore.ieee.org/document/10419367

[6] Y. Jiang, W. Zhang, and Y. Chen, “Data quality detection mechanism against label flipping attacks in federated learning,” *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1625–1637, 2023. Available: https://ieeexplore.ieee.org/document/10054157

5.9 Federated Learning Based Intrusion Detection Systems in IoT applications: An Energy Perspective

Authors: Nazlı Tekin[1]

[1] Department of Software Engineering, Erciyes University, 38280 Kayseri, Türkiye

5.9.1 Introduction

The sixth generation (6G) technology will revolutionize internet of things (IoT) applications by unfolding a fully autonomous, immersive ecosystem and enabling real-time decision-making, transforming industries and daily life like never before. 6G is set to dramatically increase the scale and efficiency of IoT applications, enabling seamless connectivity among billions of smart devices across various domains, including industrial automation, smart cities, healthcare, and smart homes. By 2030, the number of connected IoT devices is expected to exceed 40 billion worldwide [1].

IoT applications are inherently vulnerable to various security threats due to its large-scale deployment, resource-constrained devices, and diverse communication protocols. IoT applications face critical security challenges such as unauthorized access, data breaches, distributed denial-of-service (DDoS) attacks, and malware infections. Many IoT devices lack robust authentication mechanisms and are often deployed with default credentials, making them easy targets for attackers. Additionally, constrained computing resources limit the implementation of strong cryptographic measures, leaving communication channels susceptible to eavesdropping and data manipulation. To mitigate these risks, machine learning (ML)-based intrusion detection systems (IDS) are proposed to monitor network traffic, detect anomalies, and respond to potential threats in real time. However, conventional centralized ML approaches encounter fundamental constraints due to their reliance on data centralization. The necessity of aggregating data in a central location raises critical concerns regarding privacy, security, and latency.

Federated learning (FL) has emerged as a promising alternative to centralized ML, providing a decentralized approach that aligns well with the distributed and heterogeneous characteristics of IDS in IoT applications. In FL, the training process is carried out across multiple IoT devices, enabling each device to update a local model independently using its own data. These locally trained models are then aggregated at a central server to construct a global model, eliminating the need to share raw data among IoT devices. This framework effectively mitigates privacy and security concerns by ensuring data remains localized while also significantly reducing the communication overhead associated with data transmission, making it an attractive solution for IoT applications. FL approaches also support compliance with regulatory frameworks such as the general data protection regulation (GDPR), which impose strict restrictions on data transfer and storage. By minimizing the need for raw data transmission, FL serves not only as a technological advancements but also as a compliance enabler, ensuring adherence to data privacy regulations [2].

Energy efficiency is a crucial consideration in IoT applications, where numerous devices rely on constrained power sources, such as batteries or energy-harvesting mechanisms. The iterative process of FL, which requires multiple rounds of local training and communication, can impose a significant energy burden on these IoT devices, leading to rapid depletion of their power reserves. This challenge not only restricts the participation of energy-constrained IoT devices but also increases the likelihood of device dropout, potentially disrupting the FL process and diminishing the representativeness of the global model.

5.9.2 State of the Art

FL-based IDS for IoT applications play a crucial role in detecting security threats; however, their implementation on resource-constrained IoT devices poses significant challenges in energy efficiency. Most IoT devices operate on battery power, which limits their ability to perform continuous IDS. Conventional ML-based IDS approaches shift training to remote servers, introducing high latency and privacy concerns, making FL-based approaches more appealing.

Energy-Intensive Processes in FL-based IDS

FL mitigates privacy risks by enabling distributed training across IoT devices, however, its iterative nature results in frequent local model training and communication rounds, leading to excessive energy consumption. The primary contributors to energy overhead in FL-based IDS include: local model training, communication costs and security overhead. Below, we provide more details on each of these contributors.

Local model training Training ML models on low-power IoT devices requires substantial CPU and memory resources, leading to high processing power consumption and accelerated battery depletion. The complexity and size of the ML models deployed on IoT devices are major contributors to energy drain. More complex models such as convolutional neural networks (CNNs) require longer processing times, increasing energy consumption, and higher memory demands, further draining power due to frequent memory accesses and data transfers [3].

Furthermore, the characteristics of the local training dataset, such as its size, distribution, dimensionality, and preprocessing requirements, are critical determinants of energy demand in FL for IoT. Effective data-aware training strategies, such as adaptive batch sizing, sample selection, and dimensionality reduction, can help mitigate energy costs while maintaining FL-based IDS performance.

Communication costs FL requires frequent exchanges of model updates between IoT devices and the remote server to aggregate the global model. Therefore, the frequency and size of communication between IoT devices and the central aggregator significantly impact energy consumption. Additionally, the greater the number of IoT devices in an FL system, the more training and aggregation rounds occur, increasing communication overhead. Higher bandwidth demand necessitates greater transmission power, leading to higher energy consumption. Moreover, in wide-area IoT networks (e.g., smart cities, industrial IoT), devices located far from edge servers or aggregators require stronger transmission signals, further intensifying energy usage. Finally, security mechanisms to securely share the model updates (e.g., gradients or weights) can introduce additional energy overhead. Techniques such as differential privacy, secure aggregation, and encryption protocols demand extra computation and often increase the payload size of updates. Although they improve data protection and regulatory compliance, these techniques can exacerbate the energy constraints of already limited devices.

Security overhead To protect sensitive model updates, privacy-preserving techniques such as differential privacy, secure aggregation, and encryption are often employed. While these mechanisms enhance data protection and regulatory compliance, they introduce additional computational and communication overhead, which further strains energy-limited devices.

Among these factors, the most critical parameters influencing energy consumption in FL-based IDS are the model complexity, communication frequency, and use of privacy-preserving mechanisms. Moreover, many of these parameters are interdependent; for example, increasing local training to reduce communication may shift the energy burden to computation. Similarly, applying differential privacy may make the model more secure, but it increases both computation and communication costs. A comprehensive optimization strategy is therefore needed to balance these trade-offs effectively.

Optimization Techniques

To address the energy challenges of FL-based IDS, researchers have proposed various optimization techniques. Compression techniques such as pruning, quantization, and knowledge distillation effectively reduce the complexity and size of ML models, enabling execution on resource-constrained IoT devices. Pruning eliminates redundant parameters and connections, reducing model size and communication overhead in FL. Further, quantization can be applied to reduce the precision of model parameters, decreasing the memory and processing requirements. Thus, it lowers computation costs during local training on IoT devices. Additionally, knowledge distillation extracts essential features from a large global model to create a smaller, more efficient model suitable for IoT devices [3, 4].

High communication costs pose a significant challenge for many FL deployments, particularly when training large deep neural networks (DNNs), where extensive model updates are exchanged between clients and the aggregator. To this end, numerous studies have focused on enhancing communication efficiency in FL by reducing transmission frequency. Techniques such as gradient sparsification, which decreases the number of model gradients, sparse communication, where only essential local updates are transmitted, and Federated Dropout, which selectively updates a subset of parameters, help minimize energy consumption during the training process [5]. Additionally, normalization techniques such as gradient clipping, which limits the range of gradient values to prevent instability, and sign-based normalization (SignSGD), which reduces data precision by transmitting only the sign of gradients instead of full floating-point values, can be applied to encode and compress model updates.

Lightweight and TinyML Approaches

Recent research efforts have focused on Tiny Machine Learning (TinyML), which enables ultra-low-power and memory-limited IoT devices, such as microcontroller units (i.e., MCUs) to perform optimized ML models [6]. By integrating TinyML, IoT devices can locally process data and execute ML-based IDS while minimizing energy consumption and reducing reliance on cloud computing. The ability to analyze data at the extreme edge enhances real-time decision-making, enabling IoT devices to detect security threats autonomously without excessive power usage or communication overhead.

Industry leaders and researchers have recognized the potential of TinyML, leading to the development of frameworks such as TensorFlow Lite, embedded learning library (ELL) by Microsoft, ARM-NN, and STM32Cube-AI. These tools are designed to optimize ML models for deployment on microcontrollers, allowing IoT devices to run IDS locally while maintaining energy efficiency. Research evaluating the performance of tools like STM32Cube-AI and TensorFlow Lite has demonstrated that quantization techniques significantly reduce energy and memory consumption, making ML-based IDS viable for IoT networks with severe resource constraints.

Building on TinyML, tiny FL (TinyFL) has emerged as a key paradigm that extends FL to ultra-low-power edge devices. Unlike conventional FL, which often requires substantial computational and communication resources, TinyFL is designed to accommodate the extreme resource limitations of IoT devices by optimizing training architectures and communication strategies. By sharing only model weights or gradients instead of raw data, TinyFL minimizes communication costs, reduces privacy risks, and lowers energy consumption, making it a promising solution for IDS for IoT applications [7].

5.9.3 Future Directions

Future research should focus on optimizing the balance between security and energy efficiency in FL-based IDS for IoT applications.

Privacy-preserving and energy-aware FL Techniques such as differential privacy and secure aggregation enhance security but introduce computational overhead—balancing energy efficiency and security remains an open problem.

Adaptive resource management and scheduling Future work should focus on developing energy-aware resource allocation mechanisms specifically optimized for federated learning in heterogeneous IoT environments. This includes dynamic scheduling algorithms that adapt to real-time energy profiles of devices, enabling task offloading and prioritization based on residual energy levels. In addition, efficient load balancing and energy-aware client selection strategies are essential to mitigate premature device dropout and ensure consistent participation in training. Such methods are critical to maintaining the long-term operability and scalability of FL-based IDS.

TinyFL IDS in IoT applications While several TinyML frameworks exist, their integration with TinyFL remains limited. In addition, the lack of standardized frameworks and benchmarking tools to evaluate TinyML deployments presents a significant challenge to consistent evaluation of performance, scalability, and energy efficiency. Adapting these platforms and validating them through real-world federated deployments would contribute to demonstrating their effectiveness in FL-based IDS applications.

Integration of renewable energy sources Leveraging energy-harvesting techniques such as solar, kinetic, or thermal energy can enhance the long-term sustainability of FL-based IDS in energy-constrained IoT environments. These ambient energy sources enable devices to operate with reduced reliance on battery power, minimizing the need for recharging or replacements. Future FL-enabled IoT deployments may benefit from solar-powered or hybrid energy solutions to ensure uninterrupted training and inference processes.

5.9.4 References

[1] S. Sinha, “State of IoT 2024: Number of connected IoT devices growing 13% to 18.8 billion globally,” Online, 2024. Available: `https://iot-analytics.com/number-connected-iot-devices/`, Accessed at: 12-07-2025.

[2] E. Dritsas and M. Trigka, “Federated Learning for IoT: A survey of techniques, challenges, and applications,” *Journal of Sensor and Actuator Networks*, vol. 14, no. 1, p. 9, Jan. 2025. doi: 10.3390/jsan14010009

[3] N. Tekin, A. Aris, A. Acar, S. Uluagac, and V. C. Gungor, “A review of on-device machine learning for IoT: An energy perspective,” *Ad Hoc Networks*, vol. 153, p. 103348, Feb. 2024. doi: 10.1016/j.adhoc.2023.103348

[4] D. Thakur, A. Guzzo, and G. Fortino, “Hardware-algorithm co-design of energy efficient federated learning in quantized neural network,” *Internet of Things*, vol. 26, p. 101223, Jul. 2024. doi: 10.1016/j.iot.2024.101223

[5] H. Wang, L. Muñoz-González, M. Z. Hameed, D. Eklund, and S. Raza, “SparSFA: Towards robust and communication-efficient peer-to-peer federated learning,” *Computers & Security*, vol. 129, p. 103182, Jun. 2023. doi: 10.1016/j.cose.2023.103182

[6] M. Ficco, A. Guerriero, E. Milite, F. Palmieri, R. Pietrantuono, and S. Russo, “Federated learning for IoT devices: Enhancing TinyML with on-board training,” *Information Fusion*, vol. 104, p. 102189, Apr. 2024. doi: 10.1016/j.inffus.2023.102189

[7] C. N. da Silva and C. V. S. Prazeres, “Tiny federated learning for constrained sensors: A systematic literature review,” *IEEE Sensors Reviews*, vol. 2, pp. 17–31, 2025. doi: 10.1109/SR.2025.3548547

Chapter 6

Conclusions and Future Directions

As communication systems continue to evolve at an unprecedented pace, emerging technologies are increasingly shaping both legitimate applications and the facilitation of (cyber)criminal activities. While not intended to be exhaustive, this White Paper focuses on the study of communication infrastructures, with particular attention to emerging wireless technologies and their multiple facets, ranging from the physical layer to the application layer and services.

Specifically, a bottom-up approach has been adopted to examine foundational constructs. The document explores advances in physical communications interfaces, concepts of trust, current cyber threats, and the adoption of intelligent AI models. Machine learning models are analyzed through the lens of cybersecurity, with particular focus on Federated Learning – recognized for its inherent benefits in enhancing privacy and minimizing the transmission of sensitive user data – and Large Language Models (LLMs).

We argue that the dual perspectives of cybersecurity – namely, protection against “cyber criminals” and the safeguarding of the “cyber human” – are intrinsically linked and must be addressed in tandem. A comprehensive understanding of human-technology interaction is essential for effectively exploring the human dimension of cybersecurity, as such insight is critical for identifying and mitigating the underlying factors at play.

This White Paper is intended to serve as a preparatory document for future research activities within the BEiNG-WISE initiative. It lays the groundwork for investigations into advancements in cybersecurity, addressing cybercriminal activities enabled by modern automation and the use of AI systems, as well as advanced human-centered cybersecurity solutions that explicitly integrate human factors by design.