Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

FEDERATED LEARNING FOR PRIVACY-PRESERVING CYBERSECURITY: AN INTELLIGENT FRAMEWORK FOR REAL-TIME THREAT DETECTION IN IOT NETWORKS

Humna Tahreem Awan[*1], Amna Mushatq[2], Aqsa Saeed[3],

*[*1]M.E Information Technology Mehran University of Engineering and Technology, Pakistan*
[2]Teaching Assistant, Government College University Faisalabad
[3]ME Scholar, Mehran University engineering and technology Pakistan

[*1]humna.awan20@gmail.com, [2]amna_mushtaq81@yahoo.com, [3]aqsaghanghro79@gmail.com

[3]ORCID ID: 0009-0007-4286-9627

DOI: https://doi.org/10.5281/zenodo.22672718

Keywords
Federated Learning, Internet of Things, Privacy-Preserving Cybersecurity, Threat Detection, Intrusion Detection, Adversarial Robustness, Secure Aggregation, Differential Privacy, Explainable Artificial Intelligence, Real-Time Security

Article History
Received: 25 January 2026
Accepted: 10 March 2026
Published: 24 March 2026

Copyright @Author
Corresponding Author: *
Humna Tahreem Awan

Abstract

The growing Internet of Things (IoT) networks have put new pressures on cybersecurity due to distributed networks, heterogeneous and resource-constrained devices, constant data creation, sensitive data and increasingly complex attacks. The conventional centralized machine-learning methods are still dependent on shifting and consolidating data to central servers, making them susceptible to privacy threats, communication delays, and even critical weaknesses. While Federated Learning (FL) is also considered a paradigm for joint model training without the need to store raw data on the participating devices or edge nodes, it is important to note that FL is not necessarily private or secure, as updates to the models can be vulnerable to adversarial attacks, such as inference, poisoning, and backdoor attacks (Mothukuri et al., 2021; Zhang et al., 2022). In this qualitative study, how FL can facilitate threat detection via privacy-preserving and near-real-time techniques in heterogeneous IoT environments is discussed in detail while an intelligent conceptual framework is designed to combine dimensions that are typically studied separately in the existing literature. The study is conducted using an interpretivist documentary research perspective, in which the text is analyzed using thematic analysis and qualitative content analysis of literature extracted from peer-reviewed publications, authoritative cybersecurity reports, standards and relevant technical documentation. The synthesis highlights common themes in the literature related to privacy preservation, distributed threat detection, communication and computational efficiency, adversarial robustness, model poisoning, differential privacy, secure aggregation, non-IID data, device heterogeneity, explainability, trust, real-time responsiveness, governance and deployment constraints. Previous works show that FL can help to mitigate the reliance on centralized data collection but its utility relies on a delicate trade-off between privacy protection, resource usage, communication overhead, model robustness, interpretability, and operational latency (Kairouz et al., 2021; Mothukuri et al., 2021). Theoretically, the study connects the dots between

privacy and distributed intelligence, cybersecurity resilience and trustworthy AI, and practically puts forth a framework to align these dimensions in IoT threat-detection architectures. It stands out from other definitions by considering privacy, security, efficiency, explainability and adaptability in real time as different aspects of the same design challenge. The framework therefore offers a conceptual base for more reliable, privacy respecting, adaptive and decentralised IoT cybersecurity.

Introduction

The Internet of Things (IoT) has moved beyond being a collection of connected "sensing" devices and is now a distributed technology infrastructure that is integrated throughout smart homes, smart cities, healthcare systems, industrial production facilities, transportation systems, agriculture, and energy networks, among other cyber-physical systems. IoT devices are always sensing, communicating, processing and sharing information, which opens the door to intelligent automation and data-driven decision making. Meanwhile, the features that make IoT so valuable also are simultaneously posing a highly unusual cybersecurity landscape. IoT ecosystems integrate hardware with significant variations in capabilities, operating systems, communication protocols, security settings, and computational resources. A lot of endpoints are resource-limited, always connected, physically visible or without advanced security measures. Hence, security is not just a matter of protecting a centralized information system, but also a matter of securing the dynamic and distributed ecosystem in which large volumes of heterogeneous data are generated at geographically distributed locations (Briggs et al., 2021; Lazzarini et al., 2023). The variety and magnitude of such environments adds another layer of complexity to designing security mechanisms that are flexible and can adapt to evolving network conditions without placing high computational, energy or communication burden on the system.

The cybersecurity aspect of this growth is as important. Malware, botnets, denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks, unauthorized access and spoofing, ransomware, data manipulation, intrusion attempts and more and more targeted and unknown attacks are all potential targets of IoT infrastructures. Because of the distributed nature of IoT, the compromise of seemingly mundane devices could lead to a way to enter into larger networks or cyber-physical processes. It is important to have intelligent intrusion detection in place, as the static security rules may not be adequate if attacks evolve quickly and vary between devices and operating situations. Recent studies on IoT intrusion detection reveal a trend toward the use of machine learning and deep learning technologies. These methods have the potential to recognize intricate patterns in network operations and could detect unusual behavior not covered by any known signatures (Lazzarini et al., 2023). However, when new learning-based security mechanisms are applied, a new fundamental question arises – where is the data needed for intelligent learning to be processed?

Traditional machine learning architectures involve gathering data from many endpoints and sending it to a central server or data repository for storage, pre-processing, model training and learning. While there are important benefits to centralization such as relatively simple model management, aggregated datasets and computational infrastructures, there are also concerns.Centralization offers several benefits, including relatively simple model management, access to aggregated datasets, and the built-in computational infrastructure; however, there are concerns. It should therefore not be considered as inherently lacking. However, it is more questionable when IoT data contains sensitive information such as personal, operational, medical, industrial information, etc. Sharing such information in a centralized repository can raise privacy concerns, data ownership issues, significant communication burden, and other consequences in the event of a breach of the

centralized infrastructure. With large-scale IoT systems, the constant data transfer can also be inefficient due to limited bandwidth, energy, storage, and computational resources of the devices. Centralized learning and data aggregation is thus one of the major challenges faced by research on IoT intrusion detection, for privacy, scalability, latency and practical deployment (Lazzarini et al., 2023; Kairouz et al., 2021).

In this context, Federated Learning (FL) has surfaced as a promising alternative to the traditional centralized learning approach. Instead of sending all raw training information to a central repository, FL allows for learning to be done locally on the participating devices/on the organizational nodes and for the model parameters or updates to be sent to an aggregation mechanism. This collaborative process enables knowledge to be added to a common model from a variety of data sources with the original data sources held nearer to where they are stored. This is one of the reasons why FL is important for IoT systems where data are widely distributed naturally and could be constrained in terms of privacy and ownership. The recent research has identified FL as one of the promising methods for IoT security as it can minimize the direct reliance on centralized data collection and at the same time facilitate the collaborative intrusion detection process in distributed environments (Lazzarini et al., 2023; Zhang et al., 2021). Figure 1 contrasts the centralized and federated learning data flows relevant to IoT cybersecurity.

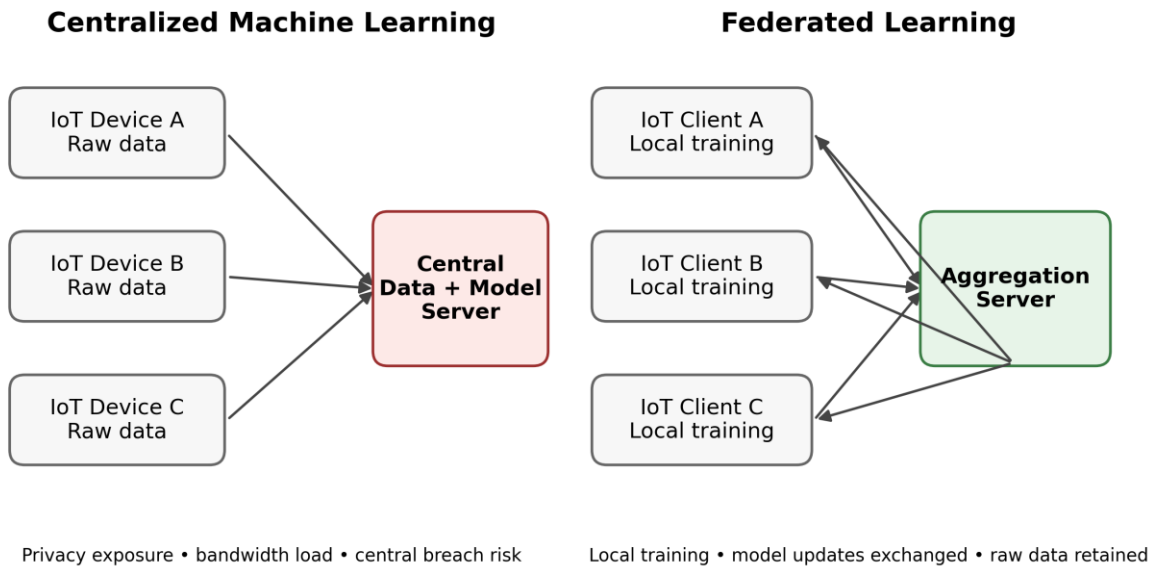


Figure 1. ***Centralized versus federated learning architecture for IoT cybersecurity.***
Note. Conceptual illustration based on Briggs et al. (2021) and Lazzarini et al. (2023).

But it is a wrong conceptualization to equate FL with privacy or security. Local storage of raw data does not mean that sensitive data cannot be derived from the data that is shared in the model updates. Furthermore, federated systems also have their own attack surfaces. Local training processes can be manipulated by malicious clients, and updates can be poisoned or backdoors can be placed in shared models. Other attacks are inference, gradient-leakage attacks, compromised communication channels, malicious participants and vulnerabilities of aggregation mechanisms. Mothukuri et al. (2021) note that security and privacy risks are not only

present during the training of a FL model, but can also appear in various stages of the FL lifecycle. Likewise, the systematic review of vulnerabilities in FL shows that decentralized data storage does not necessarily result in a trustworthy learning environment (Mothukuri et al., 2021; Zhang et al., 2022). Therefore, the question of the absolute "privacy" of FL is not what matters most, but the technical, organizational, and governance framework that could enable it to effectively support privacy-preserving cybersecurity.

The challenge is more difficult when FL is used for real-time IoT threat detection. Security systems in IoT deployments will have to be able to detect anomalies quickly enough to reduce the impact, but they will also have to do so while constrained by limited energy, processing power, network bandwidth and communication dependability. More complex models may yield better analytical capacity but may not be suitable for the low power devices; increased communication may provide better model synchronization but result in higher network overhead; more effective privacy mechanisms may result in less leakage of information but higher computational and/or analytical cost. Likewise, a model that performs well on aggregation tasks does not need to be so stable towards adversarial clients, meaningful to security practitioners, or fast enough to be deployed operationally. Literature indicates that FL-based IDS is a multi-dimensional design problem that requires evaluating the various goals, such as accuracy, latency, communication efficiency, privacy, resource usage, robustness, scalability, and trust, in combination and not separately (Lazzarini et al., 2023; Kairouz et al., 2021).

This tension raises a significant research challenge. While there has been a significant body of literature growth on FL, IoT cybersecurity, intrusion detection, and privacy-preserving machine learning, these knowledge bases often evolve independently. Some studies focus mainly on the algorithmic efficiency and effectiveness, aggregation mechanisms, or attack-detection methods, while others address privacy protection mechanisms or security flaws in FL. Other existing reviews also highlight continued challenges in non-independent and identically distributed (non-IID) data, device heterogeneity, resource constraints, communication expenses, and deployment in the real world (Lazzarini et al., 2023; Li et al., 2020). Therefore, there is still a need for an integrated interpretation of the interplay between privacy, adversarial resilience, computational requirements, communication efficiency, explainability, trust and real-time responsiveness in an IoT cybersecurity architecture.

The lack of another FL-based intrusion-detection algorithm is not the only research gap. Instead, it is about the few existing conceptual integrations of the prerequisites for FL to be a reliable and practically viable cybersecurity approach. Specifically, there is a need for more literature to focus on the relationship between non-IID and heterogeneous IoT data, resource-constrained devices, adversarial FL, mechanism for privacy enhancement like differential privacy and secure aggregation, model interpretability, and the governance aspects for distributed learning systems. These issues have been recognized in recent studies, however, they tend to be studied individually as technical issues instead of being treated as an integrated element of an intelligent cybersecurity framework (Mothukuri et al., 2021; Zhang et al., 2022). This disassociation can be challenging for researchers and practitioners to understand how the individual technical protections should work together to ensure real-time privacy-aware threat detection.

In this backdrop, the present study follows a qualitative interpretivist approach to critically summarize the current state of the art and propose an intelligent conceptual framework to enable real-time threat detection in IoT networks in a privacy-preserving manner using FL. The study looks at patterns, tensions, assumptions and relationships that emerge in existing scholarly and technical knowledge, but does not test a new algorithm or claim empirical performance. What is Federated Learning in the context of privacy-preserving mechanisms in IoT cybersecurity? (2) What are the benefits of FL in the field of cybersecurity?What are the advantages

of FL for cybersecurity? (3) What are the challenges associated with FL? (4) What are the future directions of FL in the field of cybersecurity? (2) What are the technical, privacy, security and operational problems that limit threat detection in FL environment in IoT systems? (3) What are the factors that must be considered for reliable, real-time FL based threat detection? (4) What is the way to merge all the above into a "big picture" conceptual model about intelligent IoT cybersecurity?

The study has several contributions. It is intended to theoretically bring together the views of distributed learning, preservation of privacy, cybersecurity resilience, and reliable AI into an overall interpretation. Conceptually, it shows the benefit of qualitative documentary synthesis as a means of analyzing a rapidly changing technical field in which conceptual relations can be brought to light by the fragmented algorithmic studies. On the cybersecurity side, it redefines threat detection as a multi-dimensional problem between detection capability and other notable dimensions such as privacy, adversarial robustness, explainability, communication efficiency, and operability. On the practical side, the suggested framework offers a structured approach to think about the design and governance of FL architectures in heterogeneous IoT environments, irrespective of relying solely on decentralization to mitigate the security risks. Conceptually, its key contribution is that it sees privacy, security, efficiency, trust and real-time adaptability as interconnected elements of the intelligent IoT cybersecurity landscape, not independent optimisation priorities.

This argument is then developed systematically in the rest of the article. The literature review analyzes the conceptual and empirical knowledge and research on IoT cybersecurity, FL, privacy preservation, adversarial robustness, and intelligent intrusion detection. Next, the methodology is explained according to the interpretivist documentary research design and qualitative analytical procedures. The findings and discussion draw on the literature and combine them into themes that provide the basis for developing the proposed conceptual framework. Last but not least, theoretical implications, limitations and future research and empirical validation are discussed.

Literature Review

Security Solutions: IoT Cybersecurity and Intelligent Threat Detection

The cybersecurity literature is increasingly acknowledging that the Internet of Things (IoT) environment cannot be secured using security assumptions of relatively centralized computing environments. IoT ecosystems consist of a wide variety of sensors, actuators, gateways, mobile devices, industrial controllers and edge platforms, all with significantly different properties in terms of computational power, operating systems, communication protocols and security configuration. This diversity increases the attack surface while also making security policies uniform a challenging process. Intrusion detection using machine learning has hence been the subject of great interest as it is able to detect behavioral patterns and anomalies that do not necessarily match with known signatures. While it is useful for efficiently identifying known attacks and relatively easily interpretable with rules, conventional signature-based intrusion detection is less flexible to new or rapidly evolving attacks because it relies on known attack signatures. Anomaly detection and machine learning methods are more flexible but also come with their own set of challenges, such as quality of training data, false positives, computational demands, and model generalization (Lazzarini et al., 2023).

Deep-learning methods continue to enhance the analytical abilities of IoT IDS through more intricate representation learning and from network and device data. Their appeal should not be confused with universal superiority, however. More complex models might need very large training sets and significant computing power, leading to a disconnect with resource-poor IoT devices. Furthermore, typical deep-learning systems are centralized and often demand the transport or collection of critical telemetry data, thus raising privacy and communication issues. Recent studies on the security of IoT devices thus

highlight a balance between sophisticated analysis of intelligent detection and practical limitations of distributed IoT infrastructures (Lazzarini et al., 2023; Briggs et al., 2021). The literature indicates that the problem is no longer just to detect an attack by using machine learning, but to use an intelligent attack to detect an attack over distributed environments without compromising privacy, efficiency, scalability or security.

Distributed Learning Paradigm: Federated Learning

Federated Learning (FL) is a solution to overcome some of the drawbacks related to centralizing distributed data for learning. With the traditional FL architecture, all the participating clients train the model locally and send their trained model parameters or the trained model to the aggregation server, where they are aggregated and a global model is built or updated. This can then be repeated in several rounds of communication. What makes this setup so appealing to IoT is the raw data is not automatically moved to the central repository on a regular basis. This is particularly useful for distributed IoT infrastructures where data ownership, privacy considerations, bandwidth limitations, and geographical dispersion make centralized data collection challenging (Li et al., 2020; Kairouz et al., 2021).

However, scholars are increasingly questioning uncritical narratives of FL as an inherently better alternative to centralised learning. FL doesn't remove the need for coordination, communication, aggregation, or trust, it just changes the place of computing and the means of information exchange. Some federated architectures are decentralized in the data sense but not necessarily in the control sense: there can be a central aggregation server. Furthermore, local training will have variability in data volume and distribution, hardware and infrastructure, and availability because of different clients. It is clear that federated training on these devices with limited computing resources is a significant challenge, as demonstrated by Li et al. (2020) where significant discrepancy exists between the processing and communication capabilities of the participating devices, and synchronous aggregation can be especially troublesome as highlighted by Li et al. (2020). The use of FL should therefore be read as a compromise, as it can make data centralized but adds to the coordination and optimization and security issues.

Security in FL Based Cyber Security

One of the most compelling reasons to use FL in IoT cybersecurity is for privacy. IoT data can provide insights into personal behaviour, health, industrial processes, location patterns, operational characteristics, and more. Centralized collection makes it an easy target as an attack on a central repository can expose information from many devices or organizations. The problem is addressed in part by FL that allows learning to take place near the data source. Briggs et al. (2021) point out that FL is a critical approach in privacy-preserving intrusion detection since the entities involved in the training can contribute to the development of a collaborative model without regularly sharing their raw datasets.

Data localization must not be confused, however, with wide-ranging protection for privacy. The model updates may contain information about the training data in the local area, and if an adversary can eavesdrop or disrupt the federated training process, they may exploit the model updates. In view of this, FL is now being taken into account in the literature alongside other technologies that enhance privacy. Secure aggregation aims to minimize the server's ability to see any client's contribution alone, and differential privacy aims to introduce controlled noise to make it less likely for anyone to be able to infer a record from the data. Other protection mechanisms such as encryption-based mechanisms and trusted execution environments can be added on, but come with their own computational, communication, or implementation costs. An important conceptual difference is then between retaining data locally and not being able to reconstruct or infer information from the learning process. A federated architecture does not need to achieve

the latter to achieve the former (Briggs et al., 2021; Mothukuri et al., 2021).

It is particularly salient for cybersecurity applications, where the data being secured could also include signs of malicious activity. Too much privacy protection might lead to not enough visibility for effective collaborative threat detection, and not enough might leave sensitive operational information open. Privacy preserving FL, instead, is a balancing process; it's not a trade-off between centralized and decentralized learning. There is evidence in the literature that privacy-enhancing mechanisms can be used in conjunction with FL, but there are also reports that these mechanisms can add computational overhead and that they can impact the information that is available to learning and explanation mechanisms (Briggs et al., 2021).

Identify and mitigate security vulnerabilities in Federated Learning

A key conceptual paradox in the literature is the double nature of FL, both as a privacy-preserving paradigm and as a novel attack surface for cybersecurity. Due to the fact that clients upload model updates instead of raw data, the aggregation server may have less information about the way in which the update was created. This may be helpful from a privacy perspective but a hindrance from a security perspective. It is a possibility that the training dataset, or local model parameters, can be tampered with by a malicious participant prior to submitting an update. Such poisoning attacks can adversely affect the global model, lead to targeted misclassification or introduce hidden backdoors. Importance of backdoor attacks is highlighted in recent surveys (Bagdasaryan et al., 2020) due to their ability to cause the model to behave in attacker-controlled ways if certain conditions are met, while appearing normal on other inputs.

Data poisoning, model poisoning, Byzantine behaviour, malicious clients, Sybil-type participation, gradient leakage, membership inference, and model-reconstruction or inversion are some of the other types of attack to the FL threat landscape. Mothukuri et al. (2021) shows that vulnerabilities can be present in various phases of the FL lifecycle such as local training, update transmission, aggregation, and model distribution. The resulting problem is not just protecting the server, or encryption of communication. An FL system must decide on some criteria for differentiating useful from malicious contributions and not lose the attractive decentralized properties of the architecture.

Strong aggregation and anomaly detection can help minimize the impact of suspicious updates, and cryptographic and privacy-enhancing systems can help minimize information exposure. However, there are trade-offs with each of these defenses. If client data are naturally non-IID, or if there are more stringent filtering requirements, then legitimate but atypical contributions can be suppressed as well. Differential privacy might lead to less leakage, but could also impact the usefulness of the model; secure aggregation would make it harder to inspect individual updates. Hence there is no single universal defence in the literature. Rather, it helps to direct the attention to layered security architectures that can simultaneously tackle the security challenges of confidentiality, integrity, participant trustworthiness and model robustness (Mothukuri et al., 2021; Zhang et al., 2022). Figure 2 summarizes the main attack surfaces across the FL lifecycle and the corresponding layered defenses.

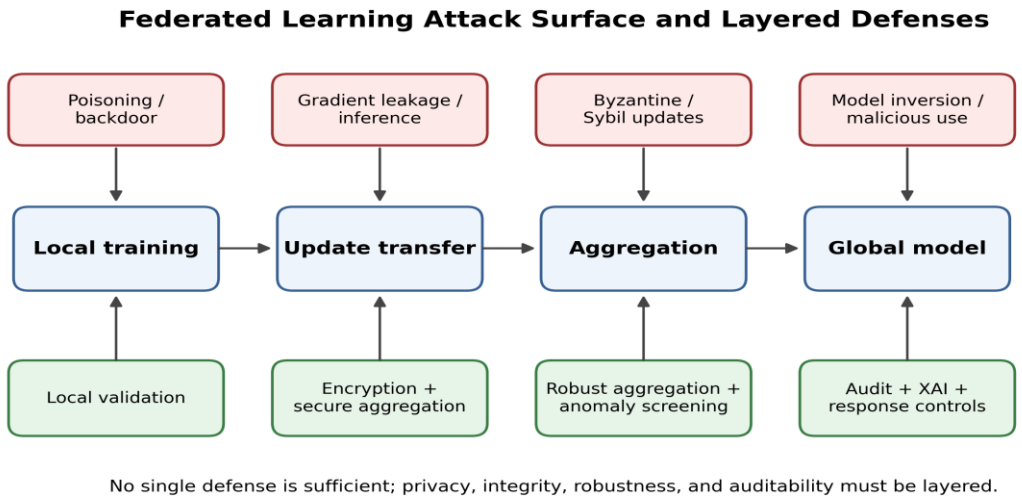


Figure 2. ***Federated learning attack surface and layered defenses.***

Note. Conceptual synthesis based on Bagdasaryan et al. (2020), Mothukuri et al. (2021), Zhang et al. (2022), and Zhu et al. (2019).

Non-IID Data and Heterogeneous IoT Environments

One of the key challenges in the real-world deployment of IoT networks is that they are heterogeneous, meaning they consist of devices with varying capabilities. Statistically identical data is seldom observed on IoT devices. A healthcare sensor and an industrial controller are used in very different environments, as are a smart-home device and an agriculture sensor. Observations may vary even for devices in the same category due to geographic location, user behaviour, hardware setups, network conditions and temporal changes. Thus, local datasets are often non-independent and non-identically distributed (non-IID).

Non-IID data may impact convergence, generalization of global models, fairness and reliability. A global model can be good for a single dominant client subset, but not so good for other, less-represented, devices or environments. Li et al. (2020) point out that data heterogeneity can have a negative impact on the accuracy of the global model and device heterogeneity introduces extra synchronization and resource utilization issues. Context can matter in attacks, too—and that's a significant topic when it comes to cybersecurity issues. What looks like an "abnormal" behaviour in one context is normal in another context.

Adding to this, there is device heterogeneity. The IoT nodes can run the complex models locally, or have very limited memory, computing power, battery life, or connectivity. Synchronous FL can impose a performance penalty on slower devices in the form of delayed aggregation, whereas withholding resources from weaker devices can give rise to representational and fairness issues. Asynchronous FL has thus become one possible solution, which poses new challenges regarding out-of-date updates, coordination and security (Li et al., 2020). However, the literature indicates that IoT FL needs to be scalable by adapting to both the heterogeneity of data and the heterogeneity of systems rather than assuming that all clients could participate equally.

In real time, detection and resource constraints

The addition of real-time cyber security adds one layer of complexity. But an intrusion-detector's worth lies as much in the ability to detect intrusions as in its ability to do so in time for meaningful action. While FL does not only

require communication rounds, local computation, aggregation and model dissemination, it also distributes computation and decreases the amount of raw-data transmitted. These processes can cause latency under bandwidth constraints or when devices are not connected for extended periods, which may not meet real-time security needs in these types of IoT scenarios. (Lazzarini et al., 2023; Kairouz et al., 2021)

Resource constraints are thus critical design factors. Li et al. (2020) show that federated learning with heterogeneous devices with limited computation presents fundamental problems. More capable (larger) models will consume more energy and local processing but could have a greater capacity for representation. Increased communication overheads but more likely to be fresh if frequent synchronization. On the other hand, fewer communication rounds can increase efficiency, but could result in slower detection of new threats. Li et al. (2020) also point to other key challenges of traditional synchronous methods, such as synchronization delays and the need for all devices to participate.

The literature thus rejects the approach to assessing IoT intrusion-detection frameworks mainly on the basis of predictive performance. Modelling and testing may show that a model can identify specific attack patterns in experiments, but this does not mean that it will be useful in practice. A practical framework should be able to simultaneously solve the issues of latency, bandwidth, energy usage, model size, communication frequency, scalability, privacy and robustness. Real-time FL-based cybersecurity is thus more of an optimization problem than a classification problem. (Kairouz et al., 2021; Li et al., 2020; Lazzarini et al., 2023) Figure 3 illustrates why real-time FL-based intrusion detection should be treated as a joint optimization problem.

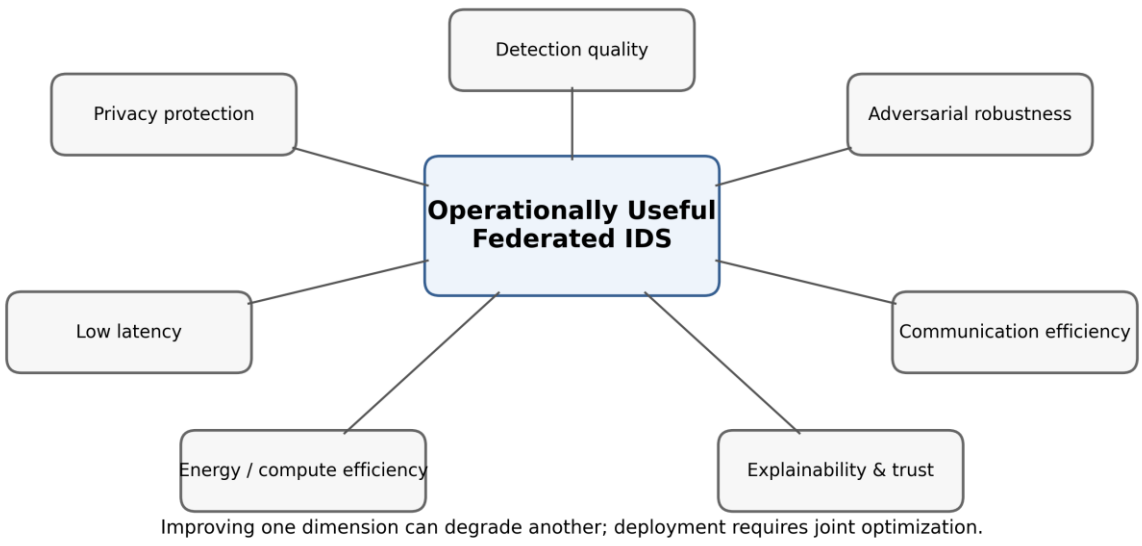


Figure 3. ***Operational trade-offs in real-time FL-based IoT threat detection.***

Note. Conceptual synthesis based on Kairouz et al. (2021), Li et al. (2020), and Lazzarini et al. (2023).

Explainable, Trustworthy, and Human-Centered FL

As AI-powered cybersecurity becomes more complex, concerns about explainability and human trust come into play. Security analysts and organizational decision makers may have to know why it was considered malicious behaviour in the system, especially when automated decisions are

made that can lead to containment measures or impact legitimate users. Explanable AI (XAI) may give insights into the characteristics, actions or trends that impact a model's decision. In FL, however, explainability gets more complex because the final FL model is based on the contributions of multiple clients whose data distributions may vary.

A model explanation that is meaningful for one client may not necessarily be meaningful for another. Variations in the local model behaviour may arise due to non-IID data, and privacy mechanisms can restrict access to information that is needed to generate detailed explanations. Direct inspection of client updates can also be prevented using secure aggregation. Based on this evidence, explainability is not an independent visualization layer that can be grafted onto FL but needs to be taken into account along with other factors such as privacy, heterogeneity, model architecture, and governance. Recent research on secure and explainable FL for IoT intrusion detection also highlights the conflict between analyst trust, privacy security, diverse explanations, and adversarial robustness (Bilal et al., 2026).

Technical model reliability is thus narrower than trust from an interpretivist view point. It covers the perception of transparency, accountability, fairness, controllability and institutional responsibility. An intelligent threat-detection system can be technically complex but hard to believe if it is not easily explainable or allow for auditing. In this context, explainability and governance are not only peripheral, but also integral parts of the social and organizational legitimacy of automated cybersecurity. (Bilal et al., 2026).

Current FL-based IoT Security Frameworks

Previous studies have shown that there are multiple approaches that could be taken for FL integration with IoT cyber security. Other approaches are mainly based on distributed intrusion detection, with local training and global aggregation to limit the telemetry to central point. For instance, Lazzarini et al. (2023) place FL as a solution to the privacy and centralization issues of traditional IoT IDS systems. Other strategies involve integrating FL with deep learning to enhance attack pattern recognition and the recent proliferation of surveys focuses on FL side-by-side with differential privacy, secure aggregation, encryption, blockchain, edge computing, and other enabling technologies (Briggs et al., 2021; Kairouz et al., 2021).

These strategies offer some positive steps forward and show signs of fragmentation. Adversarial model manipulation received comparatively less focus in frameworks that focus on privacy. Security solutions can be either poisoning based or strong aggregation solutions that do not necessarily consider the constraints of the devices or the latency of the communication. While blockchain-based solutions can enhance aspects of procedural transparency and decentralized coordination, they also have the potential to add to the computational complexity and architectural intricacies, especially in resource-limited industrial IoT environments (Kairouz et al., 2021). Likewise, models based on deep learning can offer strong representations but also may require more resources and be less interpretable.

The same tension can be seen in the literature on smart-city FL from a bigger ecosystem perspective. The implementation of smart-city infrastructures requires leveraging multiple infrastructures, stakeholders, governance frameworks, regulatory requirements and security assumptions (Kairouz et al., 2021; Briggs et al., 2021) and FL can offer a natural way to distributed learning over sensitive and heterogeneous urban data. As such, current systems tend to address specific aspects of the problem and do not offer a holistic model of the interaction between privacy, threat detection, adversarial robustness, efficiency, explainability, and real-time deployment.

Research Gap and Synthesis

Thematic synthesis is seen to show an important gap between concepts. Previous studies have already shown that FL has the potential of alleviating the need for centralized data collection systems and can be especially well-adapted for distributed IoT systems. It has also demonstrated

that FL is still susceptible to poisoning, backdoors, inference attacks, malicious participants, and other attacks. There have been investigations on non-IID data, device heterogeneity, communication efficiency, privacy enhancing technologies, and explainability via parallel scholarship. However, these dimensions are often considered individually as technical challenges. (Kairouz et al., 2021; Mothukuri et al., 2021; Lazzarini et al., 2023)

This fragmentation is an issue as the dimensions are mutually dependent. Computational cost or model utility may be impacted by stronger privacy mechanisms, interactions between non-IID data and robust aggregation, explainability may be limited by privacy, real-time detection may be at odds with frequent communication, and resource limitations may impact which security mechanisms are operationally possible. An optimization framework that focuses on optimizing one dimension, disregarding these relationships, may work well in a narrow experimental setting, but it can be challenging to implement in heterogeneous IoT environments. (Kairouz et al., 2021; Li et al., 2020; Yin et al., 2021)

The question here is not whether FL can be applied to IoT Cyber Security but how its privacy preserving properties, security weaknesses, resource constraints, varying data formats, explainability demands and real-time requirements can be "theoretically combined" into a coherent Cyber Security framework. This work fills this gap with a qualitative study where the dimensions are understood as complementary aspects of trustworthy IoT threat detection that influence each other. The study does not introduce another standalone algorithm, but rather an intelligent conceptual framework that ties the notions of privacy preservation, adversarial resilience, communication and computational efficiency, non-IID learning, explainability, trust and real-time responsiveness. It turns the literature from the question, "Does federated learning enhance intrusion detection?" – to the more critical question – under what interconnected technical and privacy, security, and governance conditions can federated intelligence be trusted and operationally useful for real-time IoT cybersecurity?

Research Methodology

Research Paradigm and Philosophical Position

The study uses interpretivist qualitative research in accordance with the nature of the research phenomenon which is knowledge construction and interpretation in the field of Federated Learning (FL), Privacy, Cyber Security, Trust and IoT Threat Detection in the field of research that has been created. Interpretivism acknowledge the fact that technological phenomena are not fully interpreted through measurable technical consequences, but also through assumptions on risk, privacy, security, accountability and practical deployment. This is especially important for FL-based cybersecurity as there are conflicting views in the literature regarding whether decentralized learning will provide better security, how security can be viewed and what is meant by 'trustworthy intelligent security'. In the absence of pre-established causal relationships, the study, instead, aims at interpreting patterns, tensions and relationships in the knowledge, and applying them to build a coherent conceptual understanding. The way is similar to qualitative research that aims at situating meaning constructions and their reflexive interaction with the data (Braun & Clarke, 2021).

Research Design

The research design used is a qualitative documentary research design using only secondary data. There is no interviewing, surveying, participant study, experiment, laboratory testing or field work involved. The design integrates systematic literature identification along with thematic synthesis to explore the scholarly and technical knowledge on FL-enabled IoT cybersecurity. Documentary research is suitable for the objective of the study because the researcher is not making any new empirical observations, but rather synthesizing an existing but dispersed body of knowledge, and doing this in a critical fashion. This study uses published research and authoritative technical documents as interpretive evidence that can be

used to examine the development, limits, assumptions, and directions of the field of FL-based cybersecurity.

Data Sources and Search Strategy

The literature was retrieved from the major scholarly databases such as IEEE Xplore, ACM Digital Library, Scopus, Web of Science, ScienceDirect, SpringerLink, Taylor & Francis, SAGE, and Google Scholar. The sources chosen were selected for their coverage of computer science, cybersecurity, artificial intelligence, information systems, IoT, and interdisciplinary computer and technology research. Where directly relevant, authoritative cybersecurity standards, institutional reports and technical documentation from recognised organisations supplemented scholarly evidence.

The search terms involved included combinations of and Boolean variations of the words "federated learning," "IoT cybersecurity," "privacy-preserving machine learning," "federated intrusion detection," "IoT threat detection," "adversarial federated learning," "secure aggregation," "differential privacy," "real-time intrusion detection," and "explainable federated learning." These terms were supplemented by other terms like non-IID data, IoT heterogeneity, model poisoning, gradient leakage, resource-constrained devices, trust, and edge computing as they came up in additional searches. Recent literature was prioritized, with a focus on studies published from 2020 to 2026, and foundational methodological and theoretical sources were kept where they were needed.

Selecting and collecting data

Relevant documents, documents that were scholarly, documents that presented their materials in an obvious method, and documents that discernibly contributed to the research problem were the four main criteria used to select documents. Studies that directly focused on FL, IoT cybersecurity, explainability, distributed learning, adversarial attacks, intrusion detection, deployment challenges, and privacy-preserving learning were given priority, while the rest were excluded. Technically relevant guidance and contextual evidence were included from authoritative standards and institutional documents. Publications that did not have a clear relevance to the research questions, provided only a short description of the methodology or concepts, or duplicated evidence without providing any new analysis, or provided only unrelated machine learning applications were excluded.

The data collection included the identification of potentially relevant documents, a review of titles and abstracts, a full-text examination of relevant documents as needed, and the organization of relevant evidence by the research questions of the study. In reading, attention was not only focused on reported abilities, but also on assumptions, limitations, contradictions, methodological choices and contextual conditions. Evidence was not reduced to numerical performance indicators, but was recorded and analysed in a thematic manner in keeping with the interpretivist orientation.

Data Analysis

Thematic analysis coupled with qualitative content analysis was used in the analysis. Thematic analysis is a flexible tool which can be used to identify themes, allowing for a sensitivity to the meaning in the context (Braun & Clarke, 2021). This process included multiple passages of reading texts, creating initial codes for key text sections, comparing codes, creating higher-order categories, and refining themes on an iterative basis. The related codes were then regrouped into thematically coherent categories related to privacy protecting, decentralized threat detection, FL security issues, IoT heterogeneity, resource limitations, real-time detection, explainability and trust, governance and deployment.

The study was not a simple description but an interpretation. Comparisons were made of similar arguments across studies, and contradictory findings and competing assumptions were directly investigated. For instance, when evaluating claims that FL would increase privacy, evidence related to a gradient leakage and inference attacks was taken into account. Likewise, proposed efficiencies were

seen in terms of communication rounds, device constraints, and differential participation. This comparison process allowed for the differences between observation and theory to be clearly understood and for relationships to be identified that could lead to the conceptual framework suggested in this study.

The trustworthiness and ethical considerations are clear

The trustworthiness was enhanced by the use of source triangulation, comparison of sources in multiple scholarly databases, careful consideration of contradictory evidence, transparent documentation of interpretive decisions, and reflexive consideration of the researchers' interpretive assumptions. Peer reviewed and authoritative sources were used to strengthen the credibility and a consistent search, selection, and coding logic strengthened the dependability. Confirmability was accomplished by seeking evidence from documents, not through assumptions or inferences made by the researcher. Reflexivity was achieved by understanding that perceptions of privacy, security, and trust vary from technological to organizational settings (Braun & Clarke, 2021). Ethical issues with respect to this study mainly focus on responsible scholarship, since the study is based on publicly available documentary evidence. The primary sources are accurately represented and cited in proper APA 7th edition format and the interpretation of the primary source does not deliberately misrepresent the conclusions. There are no manipulated participants, quotations, data sets or empirical findings.

Methodological Limitations

There are a number of caveats. Documentary research relies on published evidence, so may be influenced by publication bias, database coverage and by using terminology that varies across different sources, as well as various methodological quality. Finally, cybersecurity and FL research also evolve very quickly, thus new vulnerabilities or technologies may be introduced after the literature analyzed in the study. In addition, a qualitative synthesis is not capable of quantifying detection performance or demonstrating the superiority of the proposed framework when compared to other frameworks that are in use. Its value starts instead with conceptual integration, and critical interpretation. Therefore the framework established in this study must be considered as a theoretically supported hypothesis that it must be tested empirically, experimentally and practically in the future.

Findings and Discussion

The selected scholarly literature and authoritative technical evidence were qualitatively synthesized to arrive at five interlinked themes that elucidate the conditions and contexts under which Federated Learning (FL) can support privacy-preserving cybersecurity in Internet of Things (IoT) networks. The results are not experimental values nor are they original empirical observations. Instead, they are indicative of interpretive patterns found in the literature. One main conclusion is that FL can provide a meaningful alternative to centralized learning from an architectural perspective but requires supporting elements for privacy, adversarial defence, heterogenous environments, resource management, explainability and operational responsiveness to achieve cybersecurity benefit. Recent reviews also point to FL's ability to present opportunities for privacy as well as vulnerabilities, and highlight the fact that any FL implementation is a multi-faceted security challenge, not simply the replacement of centralized machine learning. (Mothukuri et al., 2021; Zhang et al., 2022)

Theme 1: Federated Learning as a Privacy-Preserving Architecture

The first theme is around the relatively high consensus in the literature that FL can alleviate the reliance on centralized raw-data collection for IoT security. Data collected from distributed devices is required in a conventional learning architecture to be sent to a central place for training. The basic idea of FL is to train models locally by participating devices or edge nodes and

then exchange model updates, not the original raw observations.The key idea of FL is to train the models locally at participating devices/edge nodes and then send the model updates, not the raw observations. This architecture is especially applicable to the Internet of Things as security data can include personal, operational, industrial or other sensitive information. A local data retention is thus an important privacy benefit which has been recently studied in the context of intrusion detection based on FL. (Briggs et al., 2021; Lazzarini et al., 2023)

Qualitative synthesis, however, suggests that data localization does not mean total privacy protection. Storing raw data on local devices helps mitigate one type of exposure but does not prevent leakage of information via model parameters or updates. A key point NIST makes is that information may be able to be derived from federated learning updates and trained models, even if FL is considered to ensure privacy. In a federated setting, similarly, systematic research reveals that membership, reconstruction, and inference attacks are significant threats. (Mothukuri et al., 2021; Zhu et al., 2019; Geiping et al., 2020)

The following is an important interpretation difference between privacy-preserving architecture and privacy-preserving implementation. FL provides a decentralized foundation, but other mechanisms can be used to bolster the foundation. Secure aggregation can reduce the level of visibility of individual client contributions and differential privacy can decrease the likelihood of associating model information with individual records. Additional protection can be achieved by utilizing cryptographic mechanisms, but the number of computational and communication resources required should be taken into consideration for resource-constrained IoT environments. The synthesis thus led to the development of such a framework that regards privacy not as a property that is automatically provided by FL, but rather as a layered process.

Theme 2: Intelligent Distributed Threat Detection Under IoT Heterogeneity

The second theme is the possibility of FL to enable collaborative threat detection in distributed IoT environments. Existing research increasingly seeks intrusion detection based on FL since distributed devices do not have to continuously transmit their raw security data to the shared model. This setting could enable local attack patterns to be integrated into a wider collective knowledge of cybersecurity, which could enhance detection in a variety of network situations. In particular, recent studies are pointing towards FL as a potential architecture for distributed IoT intrusion detection. (Lazzarini et al., 2023)

However, the published studies do not validate the interpretation of distributed learning as the inevitable creation of a better global model. IoT environments are very heterogeneous, the device hardware and operating conditions vary, communication capabilities are not homogenous, data distributions differ, the behavioral patterns of devices are not uniform. Local datasets can then be non-IID – i.e., they are not samples drawn from a single underlying distribution that can be used in a realistic way. Recent research for personalized federated intrusion detection has pointed out that non-IID data can impact the convergence and stability of the approach and this is an issue in approaches that assume homogeneous clients. (Li et al., 2020; Kairouz et al., 2021)

This discovery is very important for cyber security applications. The network characteristics of an attack against an industrial controller could be significantly different than those of an attack against a smart-home device. When designing a global model too heavily skewed towards predominant environments, it may not include perhaps less common but security-relevant behaviours. The framework thus introduces the concept of coordinated intelligence as an adaptive behavior rather than just adding intelligence together. The local learning, global aggregation, and threat detection should be complemented by mechanisms that can recognize heterogeneous contributions and preserve the reliability of the learned model across various IoT contexts.

This is another take on scalability, too. Increasing the number of clients is not the only way to scale FL. As more customers come on-board, they can add more communication needs, different data centers, varying availability, and more avenues for malicious involvement. The proposed framework thus considers IoT heterogeneity as a design condition of the IoT rather than an implementation issue.

Theme 3: Privacy – Security Paradox and Adversarial Robustness

Perhaps the most significant conceptual discovery is the third theme, the privacy benefits of FL coexist with major security weaknesses. While the decentralisation of raw data can decrease some privacy risks, the federated learning process also introduces new attack surfaces. Malicious participants can alter local data, or the updates to local models, which could affect the global model. Data poisoning, model poisoning, backdoor attacks, Byzantine behaviour, malicious clients, gradient leakage, membership inference attacks, and reconstruction attacks are the primary threats reported that could impact FL. (Mothukuri et al., 2021; Bagdasaryan et al., 2020; Zhu et al., 2019)

This gives rise to the "privacy–security paradox." To some extent, the architecture has been designed to prevent the transmission of sensitive raw data, but the transmitted updates could also be used in an inference or manipulation attack. The collaborative model can be affected by a malicious client even if it doesn't have direct access to another participant's raw data. Similarly, a server receiving aggregated updates can also be threatened, if the aggregation mechanism or the clients participating in the aggregation is threatened.

The literature thus argues for a layer cake defense view. Each of these approaches can be used to decrease exposure of individual client updates, provide protection from some types of information leakage, protect communication, or identify suspicious contributions using robust aggregation or anomaly-detection methods. In recent years, differential privacy and secure multi-party computation have been some of the methods under consideration for combating these threats in the context of the IoT. None of the above said mechanisms should be read as being universally sufficient.

An important consequence of this is that cyber security needs to be integrated into FL at the architectural level, not as an extra after development of the model. The proposed framework then puts adversarial defense in front of privacy protection. This is a progression from "Does FL protect data?" to more suitable questions such as "How can FL systems protect data while maintaining the integrity and reliability of collaborative learning?"

Theme :4Detection of real-time events, constraints, explanation and trust

The fourth theme is an integration of the operational and human aspects of federated IoT cybersecurity. Evidence in the literature is steadily associating FL-based intrusion detection with the rapid (or near real-time) identification of threats, but these associations must be taken with a grain of salt. Certain recent frameworks have been specifically designed for near-real-time detection, or detection capability, or for privacy. Therefore, it is not concluded from the qualitative synthesis that FL universally provides real-time cyber security.

The main challenge is between speed and distributed coordination. FL needs local computation, communication, aggregation and model updating. The processing power, battery capacity, memory and bandwidth of IoT devices are limited. Communication costs, communication frequency and energy costs can all be raised by frequent communication. Larger models might be better suited to more complex analysis, but not necessarily to smaller endpoints. For the FL in IoT, the research clearly points out heterogeneous and resource constrained devices as major challenges for practical deployment. (Li et al., 2020; Kairouz et al., 2021)

Real-time cybersecurity must then be understood as a systems level requirement and not just a model performance attribute. Detection quality, latency, communication overhead, energy, computational complexity, privacy, and scalability

are all important considerations for an effective architecture. This is in line with the recent research work offering communication optimization and robustness mechanisms for FL-based IoT intrusion detection. (Lazzarini et al., 2023; Kairouz et al., 2021)

Explainability and trust is also a part of the same theme. Consequences of cybersecurity decisions can include blocking devices, isolation of network segments, or starting incident response procedures. Security practitioners require more than an automated classification, however; they might need to be told more human reasons why a system deems behaviour suspicious. While it is possible for ExA to help with this requirement, in FL, explanations are made difficult by distributed data, heterogeneous clients, complex models, and privacy constraints.

Trust thus needs to be viewed as a more comprehensive term that includes transparency, interpretability, accountability, reliability, and human oversight. Even if the model performs well, it may be hard to implement if the security folks don't understand or audit what it does. The proposed framework thus puts explainability in the middle of automated detection and automated security responses, making it more human-centric and human-in-the-loop in defining intelligent cybersecurity, and not equating autonomous detection with a lack of human judgment.

Theme 5 : Integrated intelligent framework for privacy preserving cybersecurity in IoT environment

The fifth theme combines the results from the previous themes into the proposed Integrated Intelligent Federated Cybersecurity Framework used in the study. It is a conceptual contribution as the framework is based on qualitative interpretation of literature and is not an experimentally validated system.

The framework starts with local data generation by IoT devices. Security relevant observations are kept in their local environment and processed via local FL training. Participating devices do not send raw data, but rather model data to an orchestrated learning process. The subsequent layer adds privacy protection, including mechanisms like differential privacy and secure aggregation, if applicable. This is a stage in which the only recognition is that FL does not ensure privacy and that protection is necessary for model updates and communication processes. (Truong et al., 2021; Yin et al., 2021)

The contributions are then pooled to generate a global model that can be used for distributed threat detection. The global model is not considered a trustworthy model, however. Thus, the adversarial-defense layer is placed around aggregation and model updating to counter-attack poisoning attacks, backdoors, malicious clients and similar threats. This is followed by explainability and trust mechanisms, which enable the interpretation of the most relevant model decisions (when technically possible) and allow human security professionals to evaluate automated outputs.

Lastly, the framework adds Real-time security response and continuous model updating. The threats in IoT change over time and a static model may increasingly not reflect the present attack pattern. While it is possible to involve continuous federated updating in supporting adaptation, update frequency is acknowledged as having to be coupled with the constraints of bandwidth, energy, computation and latency. Recent research also highlights the need for optimizing communication, robustness and scalability as key factors for practical FL based IoT IDS. (Lazzarini et al., 2023; Kairouz et al., 2021) Figure 4 consolidates these elements into the integrated intelligent federated cybersecurity framework proposed in this study.

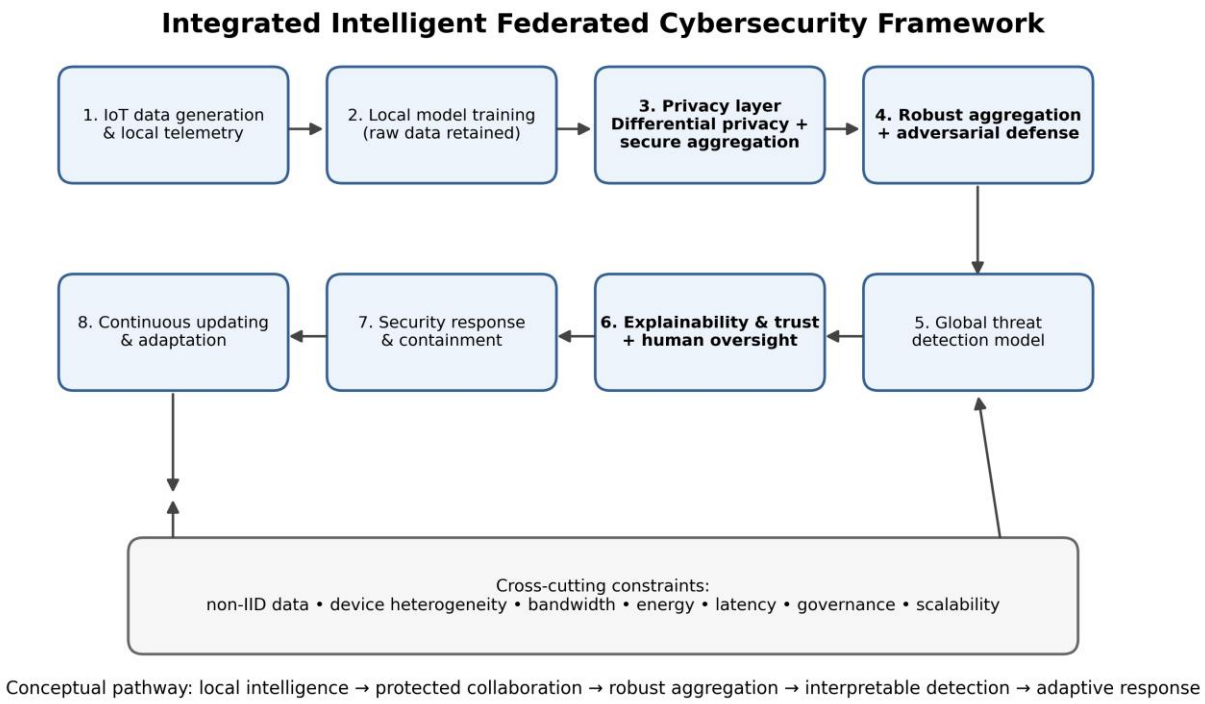


Figure 4. ***Integrated intelligent federated cybersecurity framework.***

Note. Author-developed conceptual framework synthesizing the five themes identified in this study.

The conceptual pathway that arises, in this sense, can be written as:

As the number of IoT Devices increases, the amount of data that is collected grows and it is increasingly important to identify threats to privacy and protect the data.The more IoT Devices there are, the more data is collected, and privacy protection and threat detection of the data is increasingly important.

The theoretical implications of this concept are that FL-based IoT cybersecurity should be viewed as an interdependent socio-technical architecture, rather than just a machine-learning algorithm. The impact of privacy, adversarial security, heterogeneity, resource efficiency, explainability, and responsiveness on each other. This interpretation builds on previous literature which has tended to focus on these dimensions one at a time. The recent reviews show that the threat of privacy, security issues, the diversity of IoT environments, and deployment requirements are becoming increasingly important in FL research; however, the integration of these issues into the FL paradigm is still ongoing. (Kairouz et al., 2021; Mothukuri et al., 2021)

In practice, the framework proposes that organizations refrain from assessing FL-based cybersecurity based on predictive performance. A more useful question to ask is whether the architecture is secure with regard to protecting sensitive data, malicious involvement, heterogeneous devices, resources, interpretability of output, and security responses. Most importantly, the proposed framework is not experimentally validated or guaranteed to perform in real-time. Its contribution is conceptual: it structures the conditions that have been identified in literature in a coherent way and then it can be tested for empirical and experimental research.

Overall, the results indicate that FL is a very good architectural basis for privacy sensitive IoT cybersecurity, but not without some caveats. This is because the literature has converged on the idea that there should be a holistic approach to

the following aspects of trustworthy federated cybersecurity: (i) privacy preservation, (ii) distributed intelligence, (iii) adversarial robustness, (iv) resource-aware operation, (v) explainability, and (vi) continuous adaptation. It brings together these dimensions to transcend the question of whether FL can detect attacks and address the question of how decentralized intelligence can be sufficiently trustworthy, privacy-aware, adaptive, and operationally meaningful in today's IoT cybersecurity context.

Conclusion

The Internet of Things (IoT) networks are rapidly increasing and traditional centralized learning models have begun to encounter growing conflicts between intelligent threat detection, data privacy, computational constraints, and operational responsiveness. To tackle this issue, the researchers performed a qualitative analysis by critically analyzing the current research on Federated Learning (FL), IoT cybersecurity, privacy-preserving machine learning, adversarial robustness, and trustworthy AI. The synthesis suggests that FL can serve as a valuable building block for decentralized and privacy-preserving threat detection, but does not imply privacy, security, or immunity to adversarial manipulation.

Themes that emerged from the qualitative analysis were five in number. First, FL can eliminate the need for centralized raw-data collection by allowing local model training and collaborative learning. However, data localization does not ensure privacy as model updates can be susceptible to inference and information-leakage attacks. Second, FL could enable distributed threat detection for heterogeneous IoT environments, which is affected by non-IID data, device diversity, model coordination and scalability. Third, the privacy–security paradox shows that while decentralized learning can lower some privacy risks, it also opens new attack surfaces (such as poisoning, backdoors, malicious clients and compromised aggregation). Fourth, in real-time cybersecurity, there is a need to balance detection capability, latency, communication overhead, energy consumption, computational capacity, explainability, and trust. Last, these dimensions can be embedded in an intelligent conceptual system that combines local IoT data, federated training, privacy protection, secure aggregation, global threat detection, adversarial defense, explainability, continuous updating, and security response.

The integration of dimensions which are often treated separately, therefore, constitutes the central theoretical contribution of the study. There is no such thing as privacy without cybersecurity, distributed learning without resource constraints, heterogeneity, adversarial robustness, nor communication efficiency. Just as with intelligence, predictive capability is not enough for the trustworthy deployment of a threat detection system; interpretability, accountability, and human oversight are necessary prerequisites. The proposed framework, therefore, frames FL-based cybersecurity of IoT as a socio-technical network system, instead of just a learning algorithm.

The results are also applicable. IoT organizations and cybersecurity experts need to consider FL architectures as per the full security lifecycle and not just the benefits of decentralization. Depending on the threat landscape, the system designer must take into account secure aggregation, differential privacy, robust aggregation, authentication and adversarial defenses. Technology suppliers need to build resource-efficient architectures that can run on a variety of devices and poor-quality networks. Meanwhile, policymakers should understand that alongside technology-based protection of privacy, there is a need for appropriate governance on issues of accountability, data protection, transparency and inter-organizational cooperation for privacy-preserving AI.

There are a few caveats. The study relies on documentary evidence, secondary evidence only and is unable to provide experimental evidence of detection performance or to prove that the proposed framework is superior to existing architectures. Another factor to consider is the constant evolution of FL and IoT cybersecurity; new attacks, technologies and standards may not be completely captured in the literature. The

generalizability of the qualitative interpretations is further reduced due to publication bias and context differences in research.

Research for the future should thus focus on empirical validation instead of conceptual synthesis. The proposed framework should be evaluated through experimental studies that employ benchmark datasets and realistic experimental environments for IoT, such as adversarial settings including poisoning, backdoor, inference, and privacy leakage attacks. Future studies should focus on energy-efficient FL, non-IID learning, explainable federated models, cross-device, and cross-organization deployment, and adaptation to evolving threats over time. This type of work may be used to validate the conceptual relationships found here in realistic operational conditions.

The future of IoT cybersecurity isn't the case of just installing more and more advanced artificial intelligence. It requires the development of privacy-aware, secure, explainable, resource-savvy, adaptive, and trustworthy learning architectures that can harness the technological intelligence and meet the practical demands of decentralized settings.

REFERENCES

Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). How to backdoor federated learning. Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics, 108, 2938–2948.

Bilal, M. A., Islam, I. U., Khan, M. J., Nisar, S., Farooq, M., & Khan, H. (2026). Secure and explainable federated learning for IoT intrusion detection: A comprehensive survey. IEEE Open Journal of the Communications Society, 7, 3650–3679. https://doi.org/10.1109/OJCOMS.2026.3681580

Briggs, C., Fan, Z., & Andras, P. (2021). A review of privacy-preserving federated learning for the Internet-of-Things. In M. H. ur Rehman & M. M. Gaber (Eds.), Federated learning systems: Towards next-generation AI (pp. 21–50). Springer. https://doi.org/10.1007/978-3-030-70604-3_2

Braun, V., & Clarke, V. (2021). Thematic analysis: A practical guide. SAGE.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., … Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083

Lazzarini, R., Tianfield, H., & Charissis, V. (2023). Federated learning for IoT intrusion detection. AI, 4(3), 509–530. https://doi.org/10.3390/ai4030028

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. Proceedings of Machine Learning and Systems, 2, 429–450.

Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. Future Generation Computer Systems, 115, 619–640. https://doi.org/10.1016/j.future.2020.10.007

Truong, N., Sun, K., Wang, S., Guitton, F., & Guo, Y. K. (2021). Privacy preservation in federated learning: An insightful survey from the GDPR perspective. Computers & Security, 110, Article 102402. https://doi.org/10.1016/j.cose.2021.102402

Yin, X., Zhu, Y., & Hu, J. (2021). A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions. ACM Computing Surveys, 54(6), Article 131, 1–36. https://doi.org/10.1145/3460427

Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. Knowledge-Based Systems, 216, Article 106775. https://doi.org/10.1016/j.knosys.2021.106775

Zhang, K., Song, X., Zhang, C., & Yu, S. (2022). Challenges and future directions of secure federated learning: A survey. Frontiers of Computer Science, 16, Article 165817. https://doi.org/10.1007/s11704-021-0598-z

Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. Advances in Neural Information Processing Systems, 32, 14774–14784.

Geiping, J., Bauermeister, H., Dröge, H., & Moeller, M. (2020). Inverting gradients—How easy is it to break privacy in federated learning? Advances in Neural Information Processing Systems, 33, 16937–16947.

Federated learning-based intrusion detection system for the Internet of Things using unsupervised and supervised deep learning models. (2025). Cyber Security and Applications, 3, Article 100068. https://doi.org/10.1016/j.csa.2024.100068

Federated learning for distributed intrusion detection in IoT networks. (2025). In Advanced machine learning for cyber-attack detection in IoT networks (pp. 85–110). Elsevier. https://doi.org/10.1016/B978-0-44-329032-9.00009-9

IEER
Institute for Excellence in Education & Research