Securing federated learning: a comparative review of privacy, trust, efficiency, and intrusion detection approaches

T Vamshi Krishna[1], *Anusha K*[2]

[1]Research scholar, School of Computer Science, Vellore Institute of Technology, India, Chennai
[2] Professor, School of Computer Science, Vellore Institute of Technology, India, Chennai

Abstract. Federated Learning (FL) has emerged as a privacy-preserving distributed learning paradigm that enables collaborative model training without sharing raw data. Despite its advantages, ensuring trustworthy FL deployment remains challenging due to privacy leakage risks, adversarial attacks, communication overhead, system heterogeneity, and Non-Independent and Identically Distributed (non-IID) data distributions. This review provides a comprehensive analysis of recent advancements in FL security mechanisms, including Differential Privacy (DP), secure multiparty computation, Homomorphic Encryption (HE), Byzantine-robust aggregation, blockchain-based trust integration, model compression, and post-quantum cryptographic approaches within centralized client–server architectures. The study systematically categorizes defense strategies based on their functional objectives and introduces a structured threat taxonomy to clarify attacker models and vulnerabilities. While existing works present layered security frameworks and extensive attack–defense taxonomies, most rely heavily on gradient-based optimization (e.g., Federated Averaging (FedAvg)) and lack large-scale empirical validation under realistic deployment conditions. Moreover, standardized benchmarking and multi-objective evaluation across privacy, robustness, scalability, and computational cost remain limited. This review identifies critical research gaps and emphasizes the need for integrated, deployment-aware, and empirically validated frameworks to support secure, scalable, and practical FL systems in real-world environments.

1. Introduction

Machine learning has traditionally relied on centralized architectures that aggregate data into a single repository, raising concerns related to privacy, security, scalability, and regulatory compliance [1]. The decentralized data generated through edge devices, together with strict data protection regulations, has led to the emergence of FL as a privacy-

*Corresponding author: anusha.k@vit.ac.in

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

protecting solution. Clients in FL train their models at their locations while sending only their model updates to the central aggregator, which helps them work together without exposing their complete data. FL presents multiple benefits to users, yet it brings forward additional difficulties, which include privacy leakage threats and different types of adversarial attacks, system heterogeneity, non-IID data distributions and communication challenges that occur in systems with limited resources [2]. Various mechanisms have been developed to solve these challenges, which include DP and secure aggregation methods for maintaining confidentiality, Byzantine-robust aggregation methods that protect against adversarial attacks and personalization methods that use adaptive optimization to handle non-IID data and communication-efficient methods that use compression and quantization. The existing studies mainly use gradient-based optimization methods through centralized client-server systems, which they apply to study separate elements of security, efficiency or application domains. Researchers have overlooked the development of complete systems that can unify privacy protection with system robustness, scalability and operational cost management [3].

FL allows several clients to work together on training a global model while keeping their original data protected. The clients use their confidential information to develop their models, which they would then send to the central server for model update consolidation [4]. The method decreases data exposure while helping organizations meet regulatory standards. Still, it creates problems with gradient leakage, poisoning attacks, communication overhead, and client heterogeneity, as well as problems with system stability when using non-IID data. Current research investigates secure aggregation methods together with privacy-preserving techniques and efficient communication systems and trust frameworks to enable safe deployment in real-world systems [5]. Figure 1 shows the design of FL, which enables edge devices to develop their own local models using local data and transmit their model improvements to a central aggregation server. The aggregation server processes the updates from edge devices to create a refined global model, which it sends back to the edge devices.

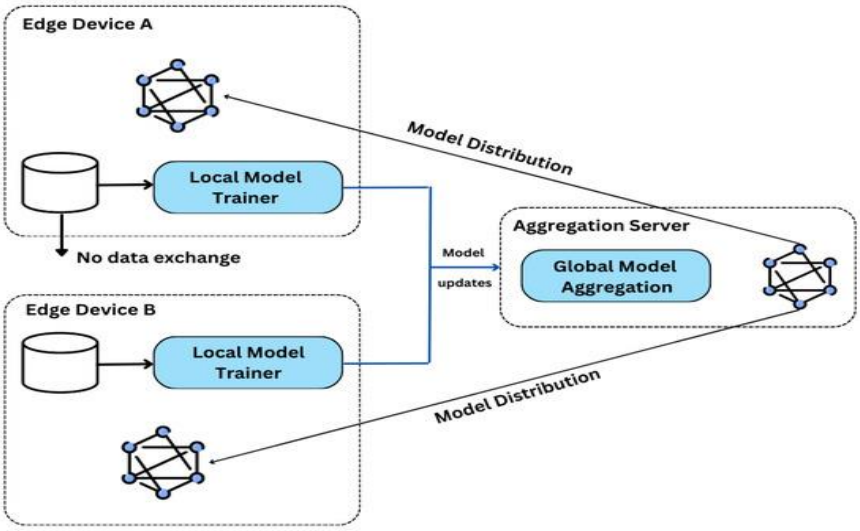

Fig 1. FL Architecture [5]

Intrusion Detection Systems (IDS) provide essential support to FL security because they identify malicious clients, unexpected updates and attacking patterns during group training. The methods used to detect security breaches in federated environments can be divided into

three distinct categories, which include model-based approaches, data-driven techniques and hybrid methods [6]. A model-based IDS which uses statistical deviation monitoring to implement rule-based filtering and data-driven techniques, which deploy machine learning models including Convolutional Neural Network (CNNs), Recurrent Neural Network (RNNs) and autoencoders, while hybrid methods use statistical learning together with deep learning models to strengthen their ability to withstand changing attack methods. The established system classification describes how IDS functions within federated security systems, while it enables structured connections between various threat categories and their corresponding detection solutions [7,8].

This review primarily focuses on gradient-based FL frameworks, particularly FedAvg, as most existing security and privacy mechanisms are designed within this optimization paradigm. Research literature currently lacks standardized assessment methods for both decentralized and non-gradient-based security approaches, which represent alternative methods to investigate security threats. The security analysis should be extended to different security models because research currently focuses only on gradient-based systems, which provide analytical stability through their main security structure. The review groups FL security mechanisms into various categories based on their main security function. The first method ensures privacy through its ability to stop all data leaks.

In contrast, the second method protects data security through its use of aggregation techniques, which can withstand both poisoning and backdoor attacks. Trust frameworks enhance auditability, while communication-efficient strategies reduce system load, and IDS-based methods are able to detect malicious clients. The analysis framework uses main goals to create groups of techniques that can achieve multiple objectives.

The study analyzes FL security through its examination of basic mechanisms that safeguard private information and defend against security breaches, establish user trust, support reliable communication and detect security incidents. The study presents three main contributions, which include a complete defense strategy classification system, a comprehensive threat evaluation framework and an implementation assessment method that evaluates system performance under different usage conditions. The research identifies essential deficiencies in current methods and demonstrates that successful secure FL implementation requires complete functional systems that maintain safety throughout their entire operational period.

2. Review Methodology

The study employed a structured literature review method to establish both methodological transparency and results that others could replicate. The study performed an extensive academic database search, which included all major databases, such as IEEE Xplore, Scopus, Web of Science, and Google Scholar. The review examines publications that span from 2021 to 2026 and display the progression of FL security from its original state to its current status. The first keyword-based search produced approximately 400 records. The publication filter brought the total studies down from 210 to the period between 2021 and 2026, which left 95 studies after researchers removed all conference papers. The review process selected a total of 15 studies after researchers completed both relevance screening and eligibility assessment. Table 1 shows the database sources and structured keyword-based search method that researchers used to find FL security literature that met their time and language requirements.

Table 1. Searching Keywords and Databases

Database	Search Strategy (TITLE-ABS-KEY)
IEEE Xplore, Scopus, Web of Science, Google Scholar	("FL" OR "federated optimization") AND ("security" OR "privacy-preserving" OR "secure aggregation" OR "Byzantine attack" OR "poisoning attack" OR "intrusion detection" OR "non-IID" OR "robust aggregation") AND PUBYEAR > 2015 AND PUBYEAR < 2026 AND (LANGUAGE: English)

Table 2. displays the study selection process, which used specific inclusion and exclusion criteria to evaluate research that met the requirements of being recent, peer-reviewed studies with empirical evidence about FL security.

Table 2. Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion	Reason
Keywords	Studies on FL security, privacy, robustness, trust, and IDS	Non-security FL works	Outside the scope of the security focus
Literature Type	Peer-reviewed journals & major conferences	Books, theses, editorials	Lack of rigorous validation
Language	English	Non-English	Ensures comparability
Timeframe	2021–2026	Before 2021	Early works not aligned with modern FL
Validation	Empirical or analytical evaluation	Concept-only papers	Insufficient methodological depth
Domain	General or domain-specific FL security	Pure centralized ML security	Not federated context

3. Related Work

This section examines the current literature about FL security, which demonstrates important research methods and study results and reveals current developments in three areas, which include privacy protection, adversarial defense and IDS.

Batur et al. (2026) [9] created an IDS that safeguards user privacy while defending against attacks that target unmanned aerial vehicle swarms through their cross-modal FL architectural system. The CM- Byzantine-Resilient Federated (BRF)-ViT framework transformed feature data into 32×32 Gramian Angular Field representations, which Vision Transformers (VIT) use to maintain resistance against Byzantine attacks in a federate system that operates with 10 clients. The experimental results demonstrated that the model reached 97.1% accuracy on the Unmanned Aerial Vehicle Intrusion Detection System 2025 Dataset (UAVIDS-2025) dataset and 78.5% accuracy on the cyber-physical dataset while achieving a fused Area Under the Curve (AUC) of 0.993 and maintaining 89.6% accuracy

when 40% of clients acted as malicious users which resulted in better performance than traditional FedAvg-based methods.

Puviarasu et al. (2026) [10] created a lightweight neural network-based FL model that uses DP and HE to achieve real-time detection without requiring direct access to unprocessed device information. The methodology utilized the IoT Intrusion Detection Dataset, comprising 1,191,264 instances and 47 features, implemented Gaussian noise-based DP, encrypted model aggregation, and audit mechanisms for compliance. The results demonstrated an overall accuracy of 94.2% (without DP) and 93.5% with encryption, with detection latency between 90–130 ms, while maintaining high attack-wise accuracy such as 94.1% for Denial of Services (DoS), 92.5% for Distributed Denial of Service (DDoS), and 93.6% for Mirai attacks, confirming the model's effectiveness in privacy-preserving real-time IoT intrusion detection.

Ramalingam et al. (2026) [11] proposed a Hybrid FL with Generative AI (HFL-GAI) framework to enhance privacy, security, and sustainability in IoT-enabled smart environments. The methodology implemented hierarchical FL across 10 IoT devices. Results showed that Hybrid FL with Generative Artificial Intelligence (HFL-GAI) achieved 96.8% accuracy and 0.97 F1-score, reduced energy consumption to 54 J (90% sustainability efficiency), and lowered communication overhead to 420 MB, outperforming Basic FL (95.2%) and FL-DP (94.1%) while maintaining scalability (93.8% accuracy with 100 clients).

Mankotia et al. (2026) [12] proposed an FL framework to enable cross-organizational intrusion detection while preserving data privacy. The approach used horizontal FL with DP ($\varepsilon = 0.5–2.0$), weighted secure aggregation, and hybrid feature selection. It was tested on five organizations with CIC-IDS-2017 (2.8M samples), University of New South Wales Network-Based Dataset 2015 (UNSW-NB15) (300,208 samples), and Bot-IoT (3.7M samples) datasets over 20 federated rounds. The experimental results indicated that the accuracy of detection was 90.3% for CIC-IDS-2017, 89.7% for UNSW-NB15, and 92.1% for Bot-IoT, with 8-15% improvement in accuracy compared to individual learning.

Alie et al. (2025) [13] proposed FL- Blockchain-based Collaborative Intrusion Detection (BCID), a FL and blockchain-integrated framework for privacy-preserving intrusion detection in Industrial IoT environments. The model was trained across 10 edge devices using federated averaging with DP ($\sigma = 1.0$) and validated through smart contracts on a permissioned blockchain over 50 rounds, evaluated on Telemetry of Network-based Internet of Things (ToN-IoT) and Network-based Detection of Botnet Attacks on Internet of Things (N-BaIoT) datasets. The framework showed 97.3% accuracy, together with 95.9% precision and 96.2% recall. In comparison, it reduced communication overhead by 41% and achieved system convergence after 21 rounds, which delivered better performance than both centralized and standard FL methods.

Shalan et al. (2025) [14] developed a privacy-preserving system for home security that uses FL, knowledge distillation, and blockchain-based role-based access control to improve home security systems. The study applied multiple techniques to the N-BaIoT dataset using their methodology, which combined transfer learning with local fine-tuning and class-score aggregation, and Confidence-Based Voting Weight (CVW)-based weighted updates over a blockchain network. Results showed centralized accuracy of around 88–91%, while FL improved device performance significantly, with the Ecobee model increasing from 58% to 87% accuracy and overall performance reaching up to 90% accuracy and 0.91 F1-score.

Chandu et al. (2025) [15] proposed a federated IDS framework incorporating Hybrid Adaptive-Weight Aggregation (HADA), SHAP-based feature selection, and DP to enhance security in distributed IoT environments. The methodology evaluated the model on the Canadian Institute for Cybersecurity – Behaviour-Centric Cybersecurity Center – National Research Council Tabular Internet of Things Attack 2024 Dataset (CIC-BCCC-NRC

TabularIoTAttack-2024) and Edge- Industrial Internet of Things (IIoT) set dataset using 100 simulated IoT devices under non-IID settings and adversarial attacks (Projected Gradient Descent (PGD), label-flip). Results showed 85–89% detection accuracy (comparable to centralized training), retained 66–73% accuracy under PGD-10 attacks, and incurred less than 1.5 percentage-point accuracy loss at $\varepsilon = 1.0$, demonstrating strong privacy–utility balance and adversarial robustness.

Timofte et al. (2025) [16] created a modular FL framework that protects privacy during distributed IoT environment intrusion detection. The methodology used gradient clipping, Fisher-based pruning, secure aggregation, DP ($\varepsilon = 1.5$), blockchain logging, and post-quantum encryption to evaluate three datasets, which included the Canadian Institute for Cybersecurity Intrusion Detection System 2017 Dataset (CICIDS2017) and TON_IoT and NSL-KDD. The results achieved 95.2% accuracy and more than 90% overall detection performance while maintaining privacy loss below 5% and demonstrating a 23-27% reduction in communication overhead, which proved strong scalability and security capabilities.

Javeed et al. (2024) [17] developed a zero-trust horizontal FL model that combined CNN and Bidirectional Long Short-Term Memory (BiLSTM) technologies to detect IoT intrusions. The system maintained local data privacy while aggregating model updates centrally and evaluating performance on CICIDS2017 and Edge-IIoTset datasets. The results demonstrated better performance than both centralized systems and traditional FL deep learning IDS methods because of enhanced threat detection abilities and improved system scalability.

Alazab et al. (2023) [18] evaluated FL for intrusion detection using the NSL-KDD dataset and compared it with traditional centralized deep learning. A horizontal FL model with federated averaging and random client selection was implemented. The FL model achieved **98.067% accuracy** with lower loss compared to centralized deep learning, demonstrating improved privacy preservation and detection effectiveness.

Rashid et al. (2023) [19] improved intrusion detection in Industrial IoT (IIoT) networks using an FL (FL) framework to preserve data privacy. The authors implemented CNN and RNN models trained on the Edge-IIoTset dataset in a federated setting and compared them with centralized ML. The proposed FL-based IDS achieved 92.49% accuracy, closely matching the centralized ML performance of 93.92%, demonstrating effective privacy-preserving detection.

Awan et al. (2023) [20] proposed FedTrust, a deep FL -based trust management framework to identify malicious and compromised IoT nodes. The model employs a unique dataset that includes 19 trust parameters and utilizes community-based FL training to achieve reduced computational requirements. The simulation results demonstrate better performance in detection and prediction tasks when compared to current trust-based methods because the system achieved higher accuracy and precision (exact values reported as superior to baselines).

Ava (2022) [21] enhanced Zero-Trust security enforcement in cloud-native distributed systems using privacy-preserving FL. A multi-cloud environment (Amazon Web Services, Azure, Google Cloud) was simulated, where anomaly detection models were trained locally and aggregated using FedAvg without sharing raw data. Experimental results show 97% overall detection accuracy, 3% false positive rate, and response times between 130–150 ms, demonstrating effective threat detection while preserving data privacy.

Attota et al. (2021) [22] achieved better attack detection results by their method, which kept data private through its design that eliminated the need for centralized data sharing. The authors segmented Message Queuing Telemetry Transport (MQTT) network data into Bidirectional-flow, Unidirectional-flow, and Packet views to use multi-view learning and selected features through Grey Wolves Optimization while using ANNs with FL across 10

virtual IoT devices. The proposed FL method outperformed the non-FL system because it reached approximately 99% accuracy, and the system maintained strong performance against adversarial situations through its 94.17% accuracy, 93.26% precision, 93.43% recall, and 94.14% F1-score results.

Babbar et al. (2021) [23] used FL and advanced cryptographic mechanisms to enhance data privacy and intrusion detection capabilities. The methodology involved local model training at vehicles using the CICIDS2017 dataset (80% training and 20% testing), application of the Laplace mechanism for DP, Federated Buffered Aggregation (FedBuff)-based buffered aggregation, and ECC with DF-HAVAL for secure communication and hash generation. The results demonstrated that the proposed FedBuff + ECC framework achieved 98.35% accuracy, 98.47% precision, and 97.52% recall with a low False Alarm Rate of 0.02.

Table 3 presents a comparative summary of recent FL-based IDS frameworks, highlighting their objectives, core techniques, and key performance outcomes across privacy, robustness, and scalability dimensions (2021–2026).

Table 3. Summary of Literature Review

Author(s)	Objective	Techniques Used	Outcome
Batur et al. (2026) [9]	To develop Byzantine-robust UAV intrusion detection	ViT, Cross-Modal FL, ReGCA aggregation	Experimental results showed 97.1% accuracy and maintained 89.6% performance with 40% malicious clients.
Puviarasu et al. (2026) [10]	To enable real-time, secure IoT intrusion detection	FL, DP, HE	Experimental results showed 94.2% accuracy with privacy loss below 5% and real-time detection latency between 90–130 ms in IoT environments.
Ramalingam et al. (2026) [11]	To improve sustainable and secure IoT learning	Hybrid FL with Generative AI (HFL-GAI), GAN/ Variational Autoencoder (VAE), Blockchain	Experimental results showed 96.8% accuracy, reducing communication overhead (~420 MB over 90 rounds) and supporting scalability up to 100 clients.
Mankotia et al. (2026) [12]	To enable collaborative privacy-preserving IoT detection	FL, DP (ε=0.5–2.0), Secure Aggregation	Experimental results showed 90–92% accuracy with less than 2.5% privacy-induced accuracy loss and 37–43% reduction in preprocessing time.
Alie et al. (2025) [13]	To integrate blockchain with IIoT intrusion detection	FL -Blockchain (FL-BCID), DP, Hyperledger Fabric	Experimental results showed 97.3% accuracy with 41% communication overhead reduction.
Shalan et al. (2025) [14]	To enhance smart home intrusion detection	FL, Knowledge Distillation, Blockchain Role-Based Access Control (RBAC)	Experimental results showed 90% overall accuracy and improved device-level performance (58% to 87%).

Chandu et al. (2025) [15]	To improve distributed IoT federated security	Hybrid Adaptive-Weight Aggregation (HADA), SHAP feature selection, DP	Results showed 85–89% accuracy with less than 1.5% accuracy loss at ε=1.0.
Timofte et al. (2025) [16]	To design a modular privacy-preserving FL cybersecurity framework	FL, Gradient Clipping, Fisher Pruning, SMPC, DP (ε=1.5), Dilithium Encryption, Blockchain	Achieved 95.2% accuracy, privacy loss below 5%, and 23–27% communication overhead reduction compared to centralized models
Javeed et al. (2024) [17]	To implement zero-trust FL-based IDS for IoT	Horizontal FL, CNN + BiLSTM, CICIDS2017, Edge-IIoTset	Outperformed centralized and standard FL models with improved detection and scalability
Alazab et al. (2023) [18]	To evaluate privacy-preserving FL for IDS	Horizontal FL, Federated Averaging (FedAvg), Random client selection, NSL-KDD	Achieved 98.067% accuracy, lower loss than centralized deep learning
Rashid et al. (2023) [19]	To improve IIoT intrusion detection privacy	FL, CNN, RNN, FedAvg, Edge-IIoTset	92.49% accuracy (FL) vs 93.92% centralized ML
Awan et al. (2023) [20]	To identify malicious IoT nodes using trust-based FL	Deep FL, Trust dataset (19 parameters), Community-based aggregation	Demonstrated higher malicious node detection and prediction performance than baseline approaches
Ava et al. (2022) [21]	To enforce zero-trust security using FL	Privacy-preserving FL, FedAvg, Multi-cloud simulation, Anomaly detection	Achieved 97% detection accuracy, 3% false positive rate, response time 130–150 ms
Attota et al. (2021) [22]	To improve privacy-preserving IoT intrusion detection	Multi-view FL, Artificial Neural Network (ANN), Grey Wolf Optimization	Experimental results showed that the model achieved ~99% accuracy and maintained 94% performance under adversarial conditions.
Babbar et al. (2021) [23]	To secure vehicular cyber-physical systems	FL, ECC, Laplace DP, FedBuff aggregation	The framework achieved 98.35% accuracy with a 0.02 false alarm rate.

4. Comparative Analysis

The selected comparative research studies show that Attota et al. (2021) achieved their highest accuracy rate of 99.0% through their multi-View FL framework, which combined artificial neural networks and Grey Wolf Optimization (GWO) methods. The study achieved superior results through its multi-view feature segmentation method, which

improved feature diversity together with GWO-based feature selection that removed duplicate features and boosted model performance. Alazab et al. (2023) achieved excellent results with a 98.067% performance score through their Horizontal FL (FedAvg) system, which used stable aggregation and maintained balanced data distribution for processing. Among the latest research works, Alie et al. (2025) (97.3%) and Batur et al. (2026) (97.1%) have shown excellent robustness by applying Blockchain, DP, and VIT-based FL architectures, which increased resistance to adversarial and Byzantine attacks while maintaining high accuracy. Although DP adds calibrated noise to protect against gradient inversion attacks, it generally leads to a slight decrease in accuracy due to the disturbance introduced in the weight update. Conversely, the relatively lower accuracy was noticed in Chandu et al. (2025) (89.0%) and Shalan et al. (2025) (90.0%) despite applying SHAP-based explainability, Knowledge Distillation, and Blockchain techniques. The slight degradation in accuracy can be justified by the presence of extra constraints related to privacy, personalization of models, and the overhead caused by the presence of interpretability layers. In general, the hybrid approach of optimized feature selection (GWO, for example), adaptive aggregation, and FL approaches performed better in terms of predictive accuracy, while the model's emphasis on interpretability and privacy preservation performed moderately well due to the additional complexity introduced by the computational overhead. The comparative analysis of the previous studies is represented in Table 4 below, and the graphical representation is shown in Figure 2.

Table 4. Comparative Analysis

Authors	Year	Technique Used	Accuracy (%)
Batur et al. [9]	**2026**	CM-BRF-ViT FL	97.1
Puviarasu et al. [10]	**2026**	FL + Diff Privacy + HE	94.2
Ramalingam et al. [11]	**2026**	HFL-GAI (Hybrid FL + GAN/VAE)	96.8
Mankotia et al. [12]	**2026**	FedPrIDS (FL + Secure Aggregation)	92.1
Alie et al. [13]	**2025**	FL-BCID (FL + Blockchain + DP)	97.3
Shalan et al. [14]	**2025**	FL + Knowledge Distillation + Blockchain	90.0
Chandu et al. [15]	**2025**	HADA-FL + SHAP + Diff Privacy	89.0
Timofte et al. [16]	**2025**	SecFL-IoT (FL + Fisher Pruning + DP + SMPC + Dilithium)	95.2
Alazab et al. [18]	**2023**	Horizontal FL (FedAvg)	98.067
Ava Smith [21]	**2022**	Privacy-Preserving FL	97
Attota et al. [22]	**2021**	Multi-View FL + ANN + GWO	99.0

Fig. 2. Comparative Analysis graphical representation.

The FL system needs a structured threat model because its distributed operation occurs in environments that have potential security threats. The review classifies threats according to three criteria, which include attacker location through three types of attackers and their purposes, which include data leakage, model corruption and service disruption and abilities through two types of attackers who operate either passively or through active or Byzantine methods and the times when attacks happen, which include training-time and inference-time. The structured taxonomy system identifies system weaknesses that researchers can use to test their security solutions while showing areas that need research and development.

5. Discussion

The findings of this review demonstrate that recent FL security research is progressively shifting from isolated privacy-preserving techniques toward integrated, multi-layered defense frameworks that combine DP, secure aggregation, Byzantine-robust methods, blockchain-based trust management, and intelligent IDS's. The research study, which analyzed different studies conducted between 2021 and 2026, found that the earlier detection systems yielded excellent results with 98 percent accuracy, whereas the modern systems performed better because of the superior design aspects that enable them to scale in a sustainable manner and also have the capability to perform in real-time. The IDS-strengthened FL architectures with their neural network capabilities and feature engineering methods, along with their hybrid aggregation approaches, demonstrate superior protection against adversarial attacks, although their effectiveness needs to be ascertained in non-IID conditions and real-world applications with different types of clients. The discussion highlights that the DP and HE privacy-preserving approaches entail additional computation time and bandwidth overheads, although the robust aggregation approaches are prone to

performance issues when handling highly diverse data. The evidence suggests that none of the approaches have the capability to handle the issues of preserving privacy and robustness, trust, and efficiency simultaneously. The future development of FL security techniques requires the creation of complete deployment-based security frameworks, which must use standardized testing metrics to achieve equal performance results in accuracy and privacy protection and system growth and operational performance during actual Internet of Things (IoT) and edge computing environments.

6. Research Challenges

The security research for FL systems must solve multiple technical problems because the research needs to address both scalability challenges and deployment difficulties. The research requires solutions for different challenges, which must be resolved before it can be applied to real-world applications in large-scale operational environments.

- The challenge of maintaining user privacy while providing necessary system functionalities remains unsolved because DP and encryption systems create problems through their accuracy reduction and greater system resource demands.
- The research fails to establish a systematic method for measuring how different techniques affect their scalability through communication costs, cryptographic expenses and system performance decline, which occurs when more clients join the system.
- The computational requirements for robust aggregation methods increase with the size of client groups because their complexity grows at a quadratic rate, which makes them impractical for usage in extensive system operations.
- The different hardware capabilities, bandwidth availability and intermittent connectivity of clients create challenges for both secure model aggregation and stable model convergence.
- The security mechanisms do not receive adequate testing because they operate under non-IID data distributions, which create natural statistical differences that can be mistaken for malicious activities.
- Many studies rely on simulation-based validation with limited client scale, lacking real-world deployment testing under operational constraints such as client dropouts and authentication management.
- The system becomes more difficult to optimize because its different security components (privacy, robustness, blockchain, IDS) create additional system complexities, which lead to increased operational delays and higher resource needs.
- The lack of standardized benchmarking frameworks and common evaluation metrics prevents appropriate evaluation and complete assessment of the scalability and deployment viability of the proposed methods.

7. Conclusion

The research emphasizes that to attain FL, robust IDS supported by neural networks and feature engineering methods is necessary. Neural network models, including CNNs, RNNs, autoencoders, and combined models, greatly enhance anomaly and malicious client detection in distributed scenarios. Feature engineering is applied to improve classification through the optimization of input data, which results in better performance for IDS. The privacy protection methods require the implementation of federated averaging together with HE and DP for safeguarding personal information, which enables the use of these advanced features. The system still suffers from four main problems, which include

difficulties with expanding capacity, the need for system resources to maintain communication, methods to manage non-IID data and protection against adversarial attacks, despite the implementation of various enhancements. The implementation of AI-based IDSs for FL systems requires real-world execution to follow more stringent guidelines, yet the technology possesses great potential. Future research needs to create ID resource-efficient designs that can be adjusted for different operating environments in IoT devices and Edge devices, which have limited processing capabilities.

7.1 Future Scope

Hybrid models use different neural networks together with advanced feature processing methods to achieve better results in detecting threats and defending against advanced attacks. The testing process needs to assess different real-world data sets. To achieve large-scale operations, it is essential to reduce computing expenses while developing better methods for maintaining user privacy. The implementation of Explainable Artificial Intelligence (XAI) would improve both trustworthiness and transparency in the decision-making process of IDS's. The FL framework is essential for building secure and scalable FL systems that can handle various operational settings while protecting against evolving cybersecurity threats. Future research should develop and evaluate FL security mechanisms that remain robust and stable under highly non-IID data distributions.

References

1. Hasan, Md Tarek, and Sai Praveen Kudapa. "Data privacy-aware machine learning and federated learning: A framework for data security." American Journal of Interdisciplinary Studies 2, no. 03 (2021): 01-34. https://doi.org/10.63125/vj1hem03
2. Wan, Yichen, Youyang Qu, Wei Ni, Yong Xiang, Longxiang Gao, and Ekram Hossain. "Data and model poisoning backdoor attacks on wireless federated learning, and the defense mechanisms: A comprehensive survey." IEEE Communications Surveys & Tutorials 26, no. 3 (2024): 1861-1897. 10.1109/COMST.2024.3361451
3. Ali, Aitizaz, Hashim Ali, Aamir Saeed, Aftab Ahmed Khan, Ting Tin Tin, Muhammad Assam, Yazeed Yasin Ghadi, and Heba G. Mohamed. "Blockchain-powered healthcare systems: enhancing scalability and security with hybrid deep learning." Sensors 23, no. 18 (2023): 7740. https://doi.org/10.3390/s23187740
4. Zheng, Yifeng, Shangqi Lai, Yi Liu, Xingliang Yuan, Xun Yi, and Cong Wang. "Aggregation service for federated learning: An efficient, secure, and more resilient realization." IEEE Transactions on Dependable and Secure Computing 20, no. 2 (2022): 988-1001. 10.1109/TDSC.2022.3146448
5. Albshaier, Latifa, Seetah Almarri, and Abdullah Albuali. "Federated learning for cloud and edge security: A systematic review of challenges and AI opportunities." Electronics 14, no. 5 (2025): 1019. https://doi.org/10.3390/electronics14051019
6. Huang, Jia, Zhen Chen, Sheng-Zheng Liu, Hao Zhang, and Hai-Xia Long. "Improved intrusion detection based on hybrid deep learning models and federated learning." Sensors 24, no. 12 (2024): 4002. https://doi.org/10.3390/s24124002
7. Lavaur, Léo, Marc-Oliver Pahl, Yann Busnel, and Fabien Autrel. "The evolution of federated learning-based intrusion detection and mitigation: a survey." IEEE Transactions on Network and Service Management 19, no. 3 (2022): 2309-2332. 10.1109/TNSM.2022.3177512

8. Kamat, Pooja, Rekha Sugandhi, and Satish Kumar. "Data-driven bearing fault detection using a hybrid autoencoder-LSTM deep learning approach." International Journal of Modelling, Identification and Control 38, no. 1 (2021): 88-103. https://doi.org/10.1504/IJMIC.2021.122471
9. Batur Şahin, Canan. "Securing UAV Swarms with Vision Transformers: A Byzantine-Robust Federated Learning Framework for Cross-Modal Intrusion Detection." Drones 10, no. 2 (2026): 125. https://doi.org/10.3390/drones10020125
10. Puviarasu, A., and V. K. Sudha. "Enhanced IoT security: privacy-preserving federated learning model for accurate, real-time intrusion detection across devices." Ain Shams Engineering Journal 17, no. 1 (2026): 103866. https://doi.org/10.1016/j.asej.2025.103866
11. Ramalingam, Venkadeshan, Basant Kumar, Shashi Kant Gupta, Deema Mohammed Alsekait, and Diaa Salama AbdElminaam. "A hybrid federated learning framework with generative AI for privacy-preserving and sustainable security in IOT-enabled smart environments." Scientific Reports 16, no. 1 (2026): 3071. https://doi.org/10.1038/s41598-025-31769-6
12. Mankotia, Sameer, Daniel Conte de Leon, and Bhaskar P. Rimal. "FedPrIDS: Privacy-Preserving Federated Learning for Collaborative Network Intrusion Detection in IoT." Journal of Cybersecurity and Privacy 6, no. 1 (2026): 10. https://doi.org/10.3390/jcp6010010
13. Ali, Anas, Mubashar Husain, and Peter Hans. "Federated learning-enhanced blockchain framework for privacy-preserving intrusion detection in industrial iot." arXiv preprint arXiv:2505.15376 (2025). https://doi.org/10.48550/arXiv.2505.15376
14. Shalan, M., M. R. Hasan, Y. Bai, and J. Li. Enhancing Smart Home Security: Blockchain-Enabled Federated Learning with Knowledge Distillation for Intrusion Detection. Smart Cities 2025, 8, 35. https://doi.org/10.3390/smartcities8010035
15. Chandu, Gutti, Thumula Karthik, and Balbudhe Parag. "Federated learning for distributed IoT security: A privacy-preserving approach to intrusion detection." IEEE Access (2025). 10.1109/ACCESS.2025.3592481
16. Timofte, Edi Marian, Mihai Dimian, Adrian Graur, Alin Dan Potorac, Doru Balan, Ionut Croitoru, Daniel-Florin Hrițcan, and Marcel Pușcașu. "Federated learning for cybersecurity: A privacy-preserving approach." Applied Sciences 15, no. 12 (2025): 6878. https://doi.org/10.3390/app15126878
17. Javeed, Danish, Muhammad Shahid Saeed, Muhammad Adil, Prabhat Kumar, and Alireza Jolfaei. "A federated learning-based zero trust intrusion detection system for Internet of Things." Ad Hoc Networks 162 (2024): 103540. https://doi.org/10.1016/j.adhoc.2024.103540
18. Alazab, Ammar, Ansam Khraisat, Sarabjot Singh, and Tony Jan. "Enhancing privacy-preserving intrusion detection through federated learning." Electronics 12, no. 16 (2023): 3382. https://doi.org/10.3390/electronics12163382
19. Rashid, Md Mamunur, Shahriar Usman Khan, Fariha Eusufzai, Md Azharuddin Redwan, Saifur Rahman Sabuj, and Mahmoud Elsharief. "A federated learning-based approach for improving intrusion detection in industrial internet of things networks." Network 3, no. 1 (2023): 158-179. https://doi.org/10.3390/network3010008
20. Awan, Kamran Ahmad, Ikram Ud Din, Mahdi Zareei, Ahmad Almogren, Byung Seo-Kim, and Jesús Arturo Pérez-Díaz. "Securing iot with deep federated learning: A trust-based malicious node identification approach." IEEE Access 11 (2023): 58901-58914. 10.1109/ACCESS.2023.3284677

21. Ava, Smith. "Privacy-Preserving Federated Learning for Zero-Trust Security Enforcement." Available at SSRN 5891085 (2022). Ava, Smith, Privacy-Preserving Federated Learning for Zero-Trust Security Enforcement (May 07, 2022). Available at SSRN: https://ssrn.com/abstract=5891085 or http://dx.doi.org/10.2139/ssrn.5891085
22. Attota, Dinesh Chowdary, Viraaji Mothukuri, Reza M. Parizi, and Seyedamin Pouriyeh. "An ensemble multi-view federated learning intrusion detection for IoT." IEEE Access 9 (2021): 117734-117745. 10.1109/ACCESS.2021.3107337
23. Babbar, Himanshi, Shalli Rani, and Mohammad Shabaz. "Federated learning with enhanced cryptographic security for vehicular cyber-physical systems." Scientific Reports 15, no. 1 (2025): 28593. https://doi.org/10.1038/s41598-025-14341-0