International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 02 | Feb 2026 www.irjet.net p-ISSN: 2395-0072

A REVIEW OF PRIVACY-PRESERVING NETWORK INTRUSION IDENTIFICATION THROUGH FEDERATED LEARNING WITH ADAPTIVE CROSS-NODE PARAMETER FUSION

KM Shrishti Sharma[1], Mrs. Arifa Khan[2]

[1]Master of Technology, Computer Science and Engineering, Lucknow Institute of Technology, Lucknow, India
[2]Assistant Professor, Department of Computer Science and Engineering, Lucknow Institute of Technology, Lucknow, India

---

Abstract - *The rapid proliferation of distributed computing environments, cloud infrastructures, and Internet-of-Things ecosystems has significantly increased the attack surface of modern networks, necessitating robust and intelligent intrusion identification mechanisms. Traditional centralized intrusion detection systems (IDS) face critical challenges related to data privacy, regulatory compliance, and cross-organizational collaboration. Federated Learning (FL) has emerged as a promising paradigm that enables collaborative model training without raw data exchange, thereby preserving data locality and confidentiality. This review systematically examines the evolution of privacy-preserving network intrusion identification frameworks based on FL, with particular emphasis on adaptive cross-node parameter fusion strategies. The paper analyzes existing architectures, aggregation algorithms, privacy-enhancing mechanisms such as differential privacy and secure aggregation, and their impact on detection performance under non-IID data distributions. A comparative taxonomy of state-of-the-art approaches is presented, highlighting trade-offs between privacy guarantees, communication efficiency, and model robustness. Furthermore, open challenges including adversarial threats, model poisoning, scalability constraints, and benchmarking inconsistencies are critically discussed. The review concludes by outlining future research directions toward resilient, adaptive, and privacy-aware federated intrusion detection systems suitable for real-world deployment.*

Key Words: Federated Learning; Network Intrusion Detection; Privacy Preservation; Adaptive Parameter Fusion; Non-IID Data; Secure Aggregation.

1. INTRODUCTION

The exponential growth of interconnected digital infrastructures, including cloud platforms, edge computing environments, and Internet-of-Things (IoT) ecosystems, has significantly expanded the cyber-attack surface. Network intrusion detection systems (IDS) play a fundamental role in identifying malicious activities, unauthorized access, and anomalous traffic patterns within such environments. With the increasing complexity of modern network architectures, conventional security mechanisms are struggling to provide scalable, privacy-compliant, and adaptive protection. Recent advances in distributed machine learning, particularly federated learning (FL), offer a promising paradigm for collaborative yet privacy-preserving intrusion identification. This review critically examines the convergence of privacy-aware federated learning frameworks and adaptive cross-node parameter fusion strategies for network intrusion detection.

1.1 Background and Motivation

The evolution of cyber threats has transitioned from isolated attacks to highly coordinated, distributed, and polymorphic intrusions targeting enterprise and critical infrastructure networks. Machine learning (ML) and deep learning (DL) techniques have significantly enhanced IDS capabilities by enabling anomaly detection and behavioral analysis beyond static signature-based approaches (Sommer and Paxson, 2010). However, effective ML-based IDS models require large-scale, diverse datasets that often reside across multiple organizations or geographically distributed nodes. Regulatory frameworks such as the General Data Protection Regulation (GDPR) further restrict centralized data sharing, creating a tension between collaborative intelligence and privacy compliance. Consequently, there is strong motivation to develop decentralized intrusion detection paradigms that preserve data locality while enabling global threat intelligence.

1.2 Limitations of Centralized Intrusion Detection

Traditional IDS architectures predominantly rely on centralized data aggregation, where raw traffic logs from distributed sources are collected and processed in a central server. Although this architecture simplifies model training and coordination, it introduces several critical limitations. First, centralized storage increases vulnerability to data breaches and single-point failures (Garcia-Teodoro et al., 2009). Second, transmitting raw network traffic incurs significant communication overhead, particularly in large-scale IoT or edge deployments. Third, centralized learning frameworks struggle with heterogeneous data distributions, as network traffic characteristics vary across domains. Additionally, privacy risks associated with sharing sensitive packet-level information hinder cross-organizational

collaboration. These structural constraints limit scalability, resilience, and regulatory compliance, thereby motivating decentralized alternatives.

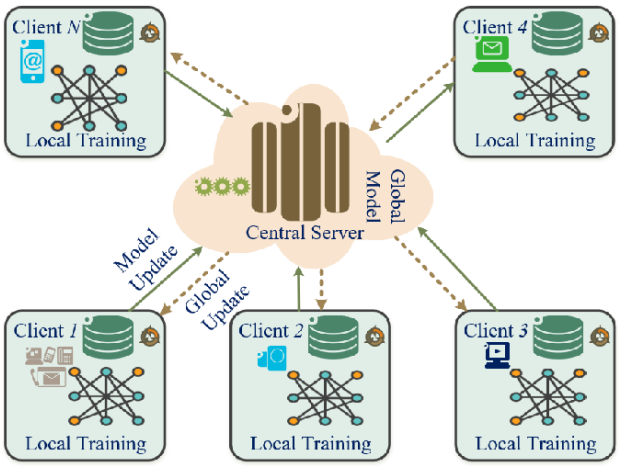

Figure-1: Federated Learning–Based IDS Architecture

1.3 Emergence of Federated Learning in Cyber security

Federated Learning, initially introduced to enable collaborative model training without centralized data exchange, has gained significant attention in privacy-sensitive domains (McMahan et al., 2017). In FL, participating nodes train local models using private datasets and share only model updates or gradients with a coordinating server. This decentralized optimization framework reduces data exposure while leveraging distributed knowledge. Within cyber security, FL has been increasingly adopted for intrusion detection, malware classification, and anomaly detection tasks, particularly in environments characterized by distributed data ownership (Nguyen et al., 2022). Despite its advantages, FL introduces new challenges, including non-independent and identically distributed (non-IID) data, communication bottlenecks, and susceptibility to adversarial model poisoning. Addressing these issues is essential for robust federated intrusion identification.

1.4 Importance of Adaptive Cross-Node Parameter Fusion

A critical component of federated learning is the parameter aggregation or fusion mechanism used to combine local model updates into a global model. Conventional algorithms such as Federated Averaging (FedAvg) apply uniform or data-size-weighted aggregation, which may be suboptimal in heterogeneous network environments (Li et al., 2020). In intrusion detection scenarios, nodes often exhibit diverse traffic patterns and threat profiles, resulting in non-IID distributions that degrade convergence and detection accuracy. Adaptive cross-node parameter fusion strategies aim to address this limitation by dynamically weighting model contributions based on trust scores, similarity metrics, data quality, or performance indicators. Such adaptive mechanisms enhance robustness against malicious participants and improve generalization across heterogeneous nodes. Therefore, adaptive fusion represents a pivotal research direction for privacy-preserving federated IDS frameworks.

2. FOUNDATIONS AND CONCEPTUAL BACKGROUND

The design of privacy-preserving federated intrusion detection systems requires an interdisciplinary understanding of network security, distributed machine learning, privacy engineering, and aggregation theory. This section provides the conceptual foundations necessary to contextualize adaptive cross-node parameter fusion within federated intrusion identification frameworks.

2.1 Network Intrusion Identification

Network intrusion identification refers to the systematic detection of malicious activities, policy violations, or anomalous behaviors within network traffic. IDS mechanisms have evolved from rule-based systems to sophisticated learning-driven architectures capable of identifying zero-day attacks and polymorphic threats.

2.1.1 Signature-Based vs Anomaly-Based IDS

Signature-based IDS operate by matching observed traffic patterns against a predefined database of known attack signatures. Systems such as Snort exemplify this approach, offering high precision for previously identified threats but limited capability against novel or obfuscated attacks (Roesch, 1999). In contrast, anomaly-based IDS establish a baseline model of normal network behavior and flag deviations as potential intrusions. This paradigm enables detection of zero-day attacks but often suffers from higher false-positive rates due to dynamic traffic patterns (Garcia-Teodoro et al., 2009). The shift toward anomaly-based detection laid the foundation for integrating machine learning techniques into IDS frameworks.

2.1.2 Machine Learning-Based IDS

Machine learning (ML) introduced statistical and algorithmic methods for modeling complex traffic behaviors. Techniques such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests demonstrated improved detection performance over traditional heuristics, particularly for high-dimensional traffic features (Bhuyan et al., 2014). ML-based IDS rely heavily on feature engineering and labeled datasets, and their effectiveness depends on the representativeness of training data. However, centralized ML training raises concerns regarding data sharing, privacy, and scalability in distributed environments.

2.1.3 Deep Learning Evolution

Deep learning (DL) further enhanced intrusion detection through automatic feature extraction and hierarchical representation learning. Architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks have demonstrated strong capability in modeling temporal dependencies and complex traffic structures (Yin et al., 2017). While DL-based IDS achieve superior accuracy, they require substantial training data and computational resources, making centralized deployment increasingly impractical in privacy-sensitive and geographically distributed infrastructures.

2.2 Federated Learning Paradigms

Federated Learning (FL) is a distributed machine learning paradigm that enables collaborative model training without sharing raw data. Instead, participating clients compute local updates that are aggregated into a global model.

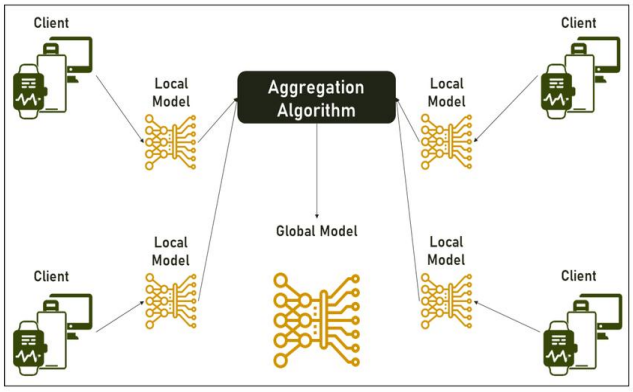

Figure-2: Federated Learning Aggregation

2.2.1 Horizontal Federated Learning

Horizontal Federated Learning (HFL) applies when participating entities share similar feature spaces but different data samples. For example, multiple organizations monitoring network traffic may collect comparable features but from distinct user bases. HFL aggregates locally trained models across these entities while maintaining data locality (McMahan et al., 2017). This paradigm is particularly suitable for cross-organizational intrusion detection.

2.2.2 Vertical Federated Learning

Vertical Federated Learning (VFL) is designed for scenarios where participants share the same sample space but possess different feature sets. In cybersecurity contexts, one entity may hold packet metadata while another maintains behavioral logs. VFL enables joint learning across complementary features without exposing sensitive attributes (Yang et al., 2019).

2.2.3 Hybrid Federated Learning

Hybrid FL integrates both horizontal and vertical settings, accommodating complex real-world environments with partially overlapping features and samples. This approach is increasingly relevant in IoT and smart infrastructure deployments where heterogeneous devices generate diverse traffic attributes.

2.2.4 Federated Optimization Algorithms

The effectiveness of FL depends heavily on its optimization strategies. Federated Averaging (FedAvg) aggregates local model parameters through weighted averaging and serves as the foundational algorithm in FL systems (McMahan et al., 2017). However, FedAvg struggles with non-independent and identically distributed (non-IID) data. FedProx introduces a proximal term to stabilize convergence under heterogeneity (Li et al., 2020), while SCAFFOLD mitigates client-drift using control variates to correct local update bias (Karimireddy et al., 2020). These optimization advancements are critical for intrusion detection, where traffic distributions vary significantly across nodes.

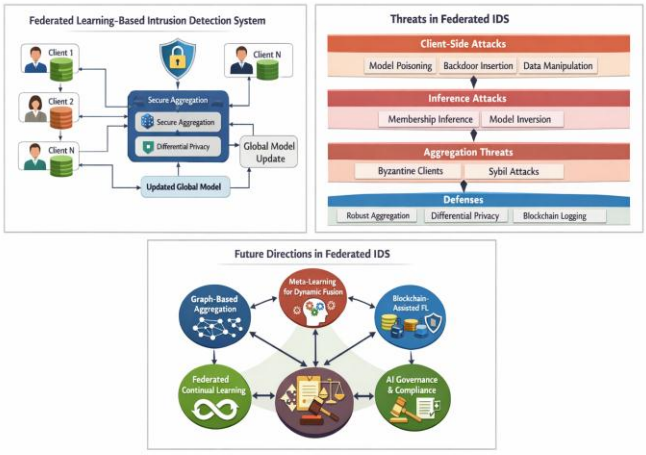

Figure-3: Federated Optimization

2.3 Privacy-Preserving Mechanisms in Federated Learning

Although FL avoids raw data sharing, model updates may still leak sensitive information. Consequently, additional privacy-enhancing mechanisms are integrated into federated frameworks.

2.3.1 Differential Privacy

Differential Privacy (DP) introduces calibrated noise into model updates to ensure that the contribution of any individual data point cannot be inferred from the aggregated model (Dwork, 2006). In federated intrusion detection, DP helps protect sensitive network traces but may degrade detection accuracy if noise levels are excessive.

2.3.2 Secure Aggregation

Secure aggregation protocols ensure that the central server can only access aggregated model parameters rather than individual client updates. Cryptographic techniques enable secure summation without revealing intermediate values (Bonawitz et al., 2017). This mechanism strengthens confidentiality in collaborative IDS environments involving semi-trusted participants.

2.3.3 Homomorphic Encryption

Homomorphic Encryption (HE) allows computations to be performed directly on encrypted data, preserving confidentiality during aggregation and optimization (Gentry, 2009). While HE provides strong privacy guarantees, its computational overhead remains a challenge for large-scale federated intrusion detection.

2.3.4 Secure Multi-Party Computation

Secure Multi-Party Computation (SMPC) enables multiple participants to jointly compute a function without revealing their individual inputs (Yao, 1982). In FL-based IDS, SMPC can support decentralized aggregation without reliance on a trusted coordinator, though scalability and latency remain concerns.

2.4 Cross-Node Parameter Fusion

Parameter fusion refers to the aggregation of locally trained models into a unified global model. In federated intrusion detection, fusion strategies directly influence convergence, robustness, and detection performance.

2.4.1 Static Aggregation

Static aggregation methods apply fixed averaging rules, typically uniform or data-size weighted averaging. While computationally efficient, static schemes assume homogeneous client behavior and may perform poorly in non-IID environments.

2.4.2 Weighted Aggregation

Weighted aggregation assigns different importance levels to clients based on criteria such as dataset size, reliability, or historical performance. This approach improves fairness and robustness compared to uniform averaging but still relies on predefined weighting schemes.

2.4.3 Similarity-Aware Fusion

Similarity-aware fusion evaluates statistical or behavioral similarity between client updates before aggregation. Distance metrics or clustering techniques identify related nodes, allowing selective aggregation to reduce the adverse effects of heterogeneous traffic distributions. Such approaches are particularly relevant in distributed intrusion detection, where network segments exhibit domain-specific patterns.

2.4.4 Adaptive and Dynamic Fusion

Adaptive fusion dynamically adjusts aggregation weights based on trust scores, performance feedback, or anomaly indicators. These methods enhance resilience against adversarial clients and model poisoning attacks by down-weighting suspicious updates. Dynamic strategies also address concept drift in evolving network traffic, making them highly suitable for real-world federated IDS deployments.

3. TAXONOMY OF FEDERATED INTRUSION DETECTION SYSTEMS

A structured taxonomy is essential to distinguish a rigorous scientific review from a descriptive survey. In the context of federated intrusion detection systems (FIDS), classification can be systematically developed along four orthogonal dimensions: learning architecture, privacy enhancement mechanism, parameter fusion strategy, and deployment environment. Such a taxonomy enables comparative analysis, highlights design trade-offs, and clarifies research gaps in privacy-preserving collaborative intrusion detection.

3.1 Taxonomy Based on Learning Architecture

Federated intrusion detection systems differ fundamentally in how coordination and communication are organized among participating nodes.

3.1.1 Centralized Coordination

In centralized federated architectures, a coordinating server orchestrates training rounds by distributing the global model to clients and aggregating local updates. This design follows the canonical federated learning model and benefits from simplified convergence control and global synchronization (Kairouz et al., 2021). For intrusion detection, centralized coordination enables consistent global threat modeling across organizations. However, it introduces partial trust assumptions toward the central aggregator and remains vulnerable to aggregation-targeted attacks or server compromise.

3.1.2 Hierarchical Federated Learning

Hierarchical federated learning introduces intermediate aggregation layers between clients and the central server. Local nodes first aggregate updates within clusters (e.g., regional networks or subnetworks) before transmitting them to a higher-level coordinator. This multi-tiered architecture reduces communication overhead and enhances scalability, particularly in IoT or large enterprise networks (Liu et al., 2020). In intrusion detection, hierarchical FL can

reflect network topology, allowing localized anomaly modeling while preserving global awareness.

3.1.3 Peer-to-Peer Federated Learning

Peer-to-peer (P2P) federated learning eliminates the need for a central coordinator by enabling decentralized model exchange among nodes. Distributed consensus or gossip-based protocols are employed to propagate updates across the network (Lian et al., 2017). In cybersecurity applications, P2P architectures enhance resilience and remove single points of failure. Nevertheless, they require robust synchronization mechanisms and are more complex to secure against adversarial participants.

3.2 Taxonomy Based on Privacy Enhancement

Although federated learning inherently avoids raw data sharing, additional mechanisms are often integrated to strengthen privacy guarantees in intrusion detection scenarios.

3.2.1 Differential Privacy-Based Systems

Differential Privacy (DP)-based federated IDS introduce calibrated noise into local gradients or model parameters before aggregation to bound information leakage (Abadi et al., 2016). This approach provides mathematically quantifiable privacy guarantees. However, privacy–utility trade-offs are significant, as excessive perturbation can degrade intrusion detection accuracy, particularly for rare attack classes.

3.2.2 Cryptographic Systems

Cryptographic approaches employ techniques such as secure aggregation and encryption-based computation to protect intermediate model updates. Secure aggregation protocols ensure that individual client contributions remain confidential even from the server (Bonawitz et al., 2017). These systems are particularly relevant for cross-organizational collaboration where mutual trust is limited. Computational complexity and communication latency, however, may limit real-time deployment in high-throughput networks.

3.2.3 Hybrid Privacy Systems

Hybrid systems combine differential privacy with cryptographic safeguards to achieve layered protection. For instance, encrypted gradient sharing may be combined with noise injection to mitigate both inference attacks and server-side leakage risks. Such multi-layered defenses aim to address sophisticated adversaries capable of exploiting model updates (Truex et al., 2019). Hybrid privacy designs are increasingly viewed as necessary for high-assurance intrusion detection environments.

3.3 Taxonomy Based on Fusion Strategy

The parameter aggregation mechanism fundamentally influences model convergence, robustness, and resilience to adversarial manipulation.

3.3.1 Uniform Averaging

Uniform averaging aggregates client updates without weighting adjustments. This strategy underpins classical federated averaging algorithms and assumes homogeneous data distributions across nodes (McMahan et al., 2017). While computationally efficient, uniform aggregation performs suboptimally in intrusion detection scenarios characterized by heterogeneous traffic patterns.

3.3.2 Data-Size Weighted Fusion

Data-size weighted fusion assigns aggregation weights proportional to the volume of local data samples. This method improves representational fairness and enhances convergence in moderately heterogeneous settings. However, it assumes that larger datasets are inherently more reliable, which may not hold if nodes are compromised or contain noisy labels.

3.3.3 Trust-Based Fusion

Trust-based aggregation incorporates reliability metrics or anomaly detection scores to weight client updates. Nodes exhibiting suspicious gradient behavior may receive reduced influence during aggregation. Such strategies mitigate model poisoning risks and improve robustness against malicious participants (Blanchard et al., 2017). In intrusion detection, trust-based fusion aligns well with adversarial threat modeling.

3.3.4 Adaptive Cross-Node Parameter Fusion

Adaptive fusion strategies dynamically adjust aggregation weights based on performance feedback, similarity measures, or statistical divergence among local models. Unlike static weighting schemes, adaptive mechanisms account for non-IID data distributions and concept drift in evolving network traffic. By leveraging similarity-aware metrics or meta-learning approaches, these systems enhance generalization and resilience in heterogeneous cybersecurity environments (Li et al., 2020). Adaptive cross-node parameter fusion thus represents a critical advancement for scalable and privacy-aware federated IDS frameworks.

3.4 Taxonomy Based on Deployment Environment

Deployment context significantly influences architectural choices, privacy requirements, and fusion strategies.

3.4.1 Cloud Environments

In cloud-based infrastructures, federated intrusion detection may operate across distributed data centers or multi-tenant platforms. Cloud environments provide substantial computational resources but require strong isolation and compliance mechanisms to prevent cross-tenant data leakage (Zissis and Lekkas, 2012). Centralized or hierarchical FL models are commonly adopted in such contexts.

3.4.2 IoT Networks

IoT networks consist of resource-constrained devices generating heterogeneous traffic patterns. Federated IDS in IoT environments must address bandwidth limitations, energy efficiency, and highly non-IID data distributions. Lightweight aggregation schemes and hierarchical coordination models are often necessary to maintain scalability.

3.4.3 Edge Computing

Edge computing environments place computation closer to data sources, reducing latency and improving responsiveness. Federated learning naturally complements edge-based IDS by enabling localized anomaly detection with periodic global synchronization (Shi et al., 2016). Edge deployments benefit from adaptive fusion strategies to handle geographically distributed traffic variations.

3.4.4 Industrial Control Systems

Industrial Control Systems (ICS) and critical infrastructure networks require highly reliable and real-time intrusion detection due to safety and operational risks. Federated IDS in ICS environments must satisfy strict latency constraints and comply with industrial cybersecurity standards. Privacy-preserving collaboration among industrial entities is particularly important to protect proprietary operational data while sharing threat intelligence.

4. COMPARATIVE LITERATURE ANALYSIS

A comparative literature analysis provides critical insight into methodological trends, empirical performance, and open limitations within federated intrusion detection systems (FIDS). Unlike descriptive surveys, this section synthesizes experimental findings across studies to evaluate architectural design, privacy enhancement techniques, aggregation strategies, and system-level efficiency. The analysis focuses on commonly reported evaluation metrics, including accuracy, F1-score, communication overhead, and robustness under heterogeneous data distributions.

4.1 Table of Key Studies

A structured comparison of key studies is typically organized using standardized evaluation parameters, including publication year, dataset utilized, federated optimization algorithm, privacy-preserving mechanism, aggregation strategy, detection accuracy, F1-score, and communication cost. Widely used benchmark datasets include NSL-KDD, UNSW-NB15, and CIC-IDS2017, each offering distinct traffic characteristics and attack diversity. For example, early federated IDS implementations primarily relied on Federated Averaging (FedAvg) evaluated on NSL-KDD, reporting competitive accuracy but limited analysis of adversarial resilience (McMahan et al., 2017). Subsequent works incorporated more realistic datasets such as UNSW-NB15 to better reflect contemporary attack patterns (Moustafa and Slay, 2015). More recent investigations have introduced privacy-aware mechanisms and heterogeneous data simulation environments, providing improved F1-scores under non-IID settings (Nguyen et al., 2022). Comparative tabulation across these studies reveals increasing attention toward communication efficiency and adaptive aggregation methods rather than purely maximizing classification accuracy.

4.2 Performance Trends and Observations

Across the literature, a consistent trend shows that federated intrusion detection models achieve performance comparable to centralized baselines when data distributions are moderately homogeneous. Deep neural network architectures, particularly CNN- and LSTM-based models, often demonstrate superior recall for complex attack classes due to enhanced feature abstraction capabilities (Yin et al., 2017). However, empirical results indicate that naive aggregation strategies may suffer from slower convergence in heterogeneous deployments. Studies incorporating proximal optimization or variance correction methods demonstrate improved stability in distributed environments (Karimireddy et al., 2020). Another observable trend is the shift from reporting only accuracy to emphasizing F1-score and recall, recognizing class imbalance as a critical issue in intrusion datasets. Furthermore, communication-efficient update compression and partial client participation have emerged as optimization priorities in large-scale deployments.

4.3 Impact of Non-IID Data on Detection Accuracy

Non-independent and identically distributed (non-IID) data significantly influence federated model convergence and generalization. In intrusion detection, traffic patterns differ across geographic locations, network roles, and organizational policies, leading to skewed class distributions. Research demonstrates that standard FedAvg experiences performance degradation under severe heterogeneity due to client drift and biased local gradients (Li et al., 2020). Empirical evaluations reveal reduced recall for minority

attack classes when data imbalance is not explicitly addressed. Techniques such as client clustering, similarity-aware aggregation, and personalized federated learning have been proposed to mitigate these effects (Hanzely and Richtárik, 2020). Overall, non-IID data remains one of the most critical challenges in achieving robust federated intrusion detection.

4.4 Privacy–Utility–Communication Trade-offs

The integration of privacy-enhancing mechanisms introduces measurable trade-offs among detection performance, communication efficiency, and confidentiality guarantees. Differential privacy, while offering formal privacy bounds, often results in reduced model precision due to gradient perturbation (Abadi et al., 2016). Secure aggregation protocols enhance confidentiality but increase computational and communication overhead, particularly in high-frequency training rounds (Bonawitz et al., 2017). Communication compression techniques, such as gradient sparsification or quantization, reduce bandwidth consumption but may slow convergence or affect minority-class detection accuracy (Konečný et al., 2016). Comparative analysis across studies indicates that achieving optimal balance requires adaptive tuning of privacy budgets, aggregation frequency, and update granularity. Consequently, privacy–utility–communication trade-offs represent a multidimensional optimization problem central to the design of scalable federated IDS frameworks.

5. SECURITY THREATS IN FEDERATED IDS

While federated learning enhances privacy by avoiding raw data centralization, it introduces a new attack surface at the model-update level. In federated intrusion detection systems (FIDS), adversaries may manipulate training dynamics, inject malicious updates, or extract sensitive information from shared parameters. Unlike traditional centralized IDS, federated environments must defend against both external cyber threats and internal adversarial participants. This section examines principal attack categories and discusses aggregation-level defenses relevant to privacy-preserving federated IDS frameworks.

5.1 Model Poisoning Attacks

Model poisoning attacks occur when malicious clients intentionally manipulate local training updates to degrade global model performance or bias predictions. In federated IDS, a compromised node may alter gradient values or inject adversarial perturbations that skew attack classification boundaries. Such attacks are particularly effective because the central server typically assumes that local updates are benign and aggregates them without direct access to raw data. Byzantine-resilient attack models demonstrate that even a small fraction of malicious participants can significantly distort convergence in distributed learning (Blanchard et al., 2017). Empirical evidence shows that poisoning can reduce detection recall for critical attack classes while maintaining overall accuracy, thereby masking degradation in intrusion detection effectiveness (Bhagoji et al., 2019). The decentralized nature of federated IDS amplifies this risk, especially in cross-organizational deployments where trust relationships are limited.

5.2 Backdoor Attacks

Backdoor attacks represent a specialized form of model poisoning in which adversaries embed hidden triggers into the global model. In intrusion detection scenarios, a malicious client may train its local model to misclassify specific attack signatures when a predefined traffic pattern appears. The global aggregation process then inadvertently incorporates this backdoor behavior. Unlike general poisoning, backdoor attacks aim to preserve overall performance while enabling targeted evasion. Studies have shown that model replacement techniques can allow adversaries to scale malicious updates to dominate aggregation rounds without detection (Bagdasaryan et al., 2020). In federated IDS environments, such attacks are particularly dangerous because they may enable stealthy bypass of security monitoring systems while remaining statistically inconspicuous.

5.3 Inference Attacks (Membership and Model Inversion)

Inference attacks exploit shared model updates to extract sensitive information about local training data. Membership inference attacks aim to determine whether specific data samples were part of a client's training set, potentially revealing confidential network traffic patterns (Shokri et al., 2017). Model inversion attacks go further by reconstructing approximations of original training inputs from gradient information (Zhu et al., 2019). In federated intrusion detection, exposure of traffic metadata or behavioral signatures through such attacks could compromise organizational privacy or reveal operational vulnerabilities. Although federated learning reduces direct data exposure, gradient leakage remains a documented risk, particularly when privacy-preserving mechanisms are not rigorously implemented.

5.4 Robust Aggregation Countermeasures

To mitigate adversarial threats, robust aggregation mechanisms have been developed to detect and neutralize malicious updates. Byzantine-resilient aggregation methods, such as Krum and trimmed-mean algorithms, attempt to filter anomalous gradients before global model updates (Blanchard et al., 2017). Median-based or norm-clipping strategies further reduce the influence of extreme parameter deviations. More advanced defenses incorporate anomaly detection at the update level, evaluating statistical divergence or trust scores before aggregation (Pillutla et al., 2022). In privacy-preserving federated IDS, combining

robust aggregation with differential privacy and secure aggregation protocols can enhance resilience against both poisoning and inference attacks. However, defensive mechanisms often introduce additional computational overhead and may slow convergence, highlighting the need for adaptive and efficient countermeasures tailored to cyber security contexts.

6. EMERGING RESEARCH DIRECTIONS

The evolution of federated intrusion detection systems (FIDS) is increasingly shaped by advances in distributed optimization, adaptive intelligence, and secure coordination mechanisms. While foundational federated learning frameworks have demonstrated feasibility in privacy-preserving intrusion detection, emerging research directions aim to enhance robustness, scalability, and regulatory alignment. This section synthesizes key forward-looking paradigms that are expected to influence the next generation of adaptive cross-node federated IDS architectures.

6.1 Graph-Based Federated Aggregation

Traditional federated aggregation assumes a star-topology communication structure where clients interact only with a central coordinator. However, real-world network infrastructures often exhibit graph-like relationships among nodes. Graph-based federated aggregation leverages graph neural networks (GNNs) or topology-aware optimization strategies to model inter-client relationships during parameter fusion. By incorporating structural similarity or communication proximity into aggregation weights, graph-based methods improve convergence under heterogeneous conditions (Scarselli et al., 2009). In intrusion detection, graph-based aggregation enables clustering of nodes with similar traffic distributions, thereby reducing the adverse effects of non-IID data. Furthermore, topology-aware aggregation can enhance resilience against malicious nodes by isolating anomalous participants within graph partitions.

6.2 Meta-Learning for Dynamic Fusion

Meta-learning, often described as "learning to learn," provides a framework for adaptive model generalization across heterogeneous environments. In federated IDS, meta-learning techniques can dynamically adjust aggregation parameters or initialization weights to optimize performance under varying traffic distributions. Model-Agnostic Meta-Learning (MAML) and related approaches allow rapid adaptation to new attack patterns with minimal local retraining (Finn et al., 2017). Integrating meta-learning into federated aggregation facilitates personalized or cluster-specific models, addressing client drift and concept evolution in distributed networks. This direction is particularly relevant for adaptive cross-node parameter fusion, where aggregation rules can themselves be optimized through meta-objectives.

6.3 Block chain-Assisted Federated IDS

Blockchain technology offers decentralized trust management and tamper-resistant record-keeping, which can complement federated learning in adversarial environments. In federated IDS, block chain can be employed to securely log model updates, verify participant integrity, and enforce smart-contract-based aggregation policies (Zhang et al., 2020). This integration reduces reliance on a fully trusted central server and enhances transparency in collaborative cyber security settings. However, block chain-assisted federated systems must address scalability constraints and consensus latency, especially in high-throughput intrusion detection scenarios. The convergence of block chain and federated learning represents a promising pathway toward decentralized, auditable threat intelligence sharing.

6.4 Federated Continual Learning

Network environments are dynamic, with evolving attack vectors and shifting traffic patterns. Federated continual learning extends standard FL by enabling incremental model updates without catastrophic forgetting of previously learned attack signatures. Continual learning mechanisms, such as regularization-based memory retention and replay strategies, have been proposed to preserve historical knowledge while incorporating new threat data (Parisi et al., 2019). In federated IDS, continual learning supports long-term adaptability across distributed nodes, allowing systems to remain responsive to emerging cyber threats. Combining continual learning with adaptive aggregation strategies can further mitigate performance degradation under concept drift conditions.

6.5 AI Governance and Regulatory Compliance

As federated intrusion detection systems are deployed across jurisdictions and industries, compliance with regulatory frameworks becomes increasingly critical. AI governance principles emphasize transparency, accountability, and explain ability in machine learning systems (Floridi et al., 2018). In cyber security applications, explainable federated IDS models can improve trust among collaborating entities and support forensic investigations. Additionally, privacy-preserving mechanisms must align with legal standards such as GDPR and sector-specific cyber security regulations. Emerging research focuses on embedding compliance-aware constraints into federated optimization processes and establishing standardized evaluation protocols for privacy guarantees. Governance-aligned federated IDS architectures are expected to play a central role in responsible and sustainable cyber security collaboration.

7. CONCLUSION

This review has systematically examined the evolving landscape of privacy-preserving network intrusion identification through federated learning with adaptive cross-node parameter fusion. By synthesizing existing research across architectural paradigms, privacy-enhancement mechanisms, aggregation strategies, and deployment contexts, the study highlights the transition from conventional centralized intrusion detection systems toward collaborative and decentralized intelligence frameworks. The analysis demonstrates that federated learning effectively mitigates raw data exposure risks while maintaining competitive detection performance. However, the success of federated intrusion detection systems (FIDS) largely depends on robust aggregation mechanisms capable of handling non-IID data, adversarial manipulation, and dynamic network conditions. Adaptive cross-node parameter fusion emerges as a critical enabler, improving convergence stability, resilience against poisoning attacks, and generalization across heterogeneous environments. Furthermore, emerging directions such as graph-based aggregation, federated continual learning, and governance-aligned AI frameworks indicate a maturation of the field toward scalable and regulation-compliant deployment. Overall, privacy-aware federated IDS represent a promising paradigm for secure collaborative cyber security, though further advancements in robustness, standardization, and real-world validation remain necessary.

7.1. Limitations of This Review

Despite providing a structured taxonomy and comparative synthesis, this review has certain limitations. First, the analysis relies on published academic literature, which may not fully capture proprietary or industry-deployed federated intrusion detection implementations. Second, variations in experimental setups, datasets, and evaluation metrics across studies limit direct quantitative comparability. Third, rapidly evolving federated optimization and adversarial defense techniques may render some observations temporally constrained. Additionally, while privacy–utility trade-offs are discussed conceptually, a formal meta-analysis was not conducted due to inconsistent reporting of communication overhead and privacy budgets. Finally, the review emphasizes adaptive cross-node parameter fusion within federated frameworks, potentially underrepresenting alternative decentralized learning paradigms that could also support privacy-preserving intrusion detection.

REFERENCES

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K. and Zhang, L. (2016) 'Deep learning with differential privacy', Proceedings of the ACM Conference on Computer and Communications Security (CCS), pp. 308–318.
2. Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D. and Shmatikov, V. (2020) 'How to backdoor federated learning', Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 2938–2948.
3. Bhagoji, A.N., Chakraborty, S., Mittal, P. and Calo, S. (2019) 'Analyzing federated learning through an adversarial lens', Proceedings of the International Conference on Machine Learning (ICML), pp. 634–643.
4. Bhuyan, M.H., Bhattacharyya, D.K. and Kalita, J.K. (2014) 'Network anomaly detection: Methods, systems and tools', IEEE Communications Surveys & Tutorials, 16(1), pp. 303–336.
5. Blanchard, P., El Mhamdi, E.M., Guerraoui, R. and Stainer, J. (2017) 'Machine learning with adversaries: Byzantine tolerant gradient descent', Advances in Neural Information Processing Systems, 30, pp. 119–129.
6. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H.B., Patel, S., Ramage, D., Segal, A. and Seth, K. (2017) 'Practical secure aggregation for privacy-preserving machine learning', Proceedings of the ACM CCS, pp. 1175–1191.
7. Dwork, C. (2006) 'Differential privacy', Proceedings of the International Colloquium on Automata, Languages and Programming (ICALP), pp. 1–12.
8. Finn, C., Abbeel, P. and Levine, S. (2017) 'Model-agnostic meta-learning for fast adaptation of deep networks', Proceedings of ICML, pp. 1126–1135.
9. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E. (2018) 'AI4People—An ethical framework for a good AI society', Minds and Machines, 28(4), pp. 689–707.
10. Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G. and Vázquez, E. (2009) 'Anomaly-based network intrusion detection: Techniques, systems and challenges', Computers & Security, 28(1–2), pp. 18–28.
11. Gentry, C. (2009) 'Fully homomorphic encryption using ideal lattices', Proceedings of the ACM Symposium on Theory of Computing (STOC), pp. 169–178.
12. Hanzely, F. and Richtárik, P. (2020) 'Federated learning of a mixture of global and local models', arXiv preprint arXiv:2002.05516.
13. 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., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B.,

Gibbons, P.B., Green, M., Harchaoui, Z., He, Y., He, T., Huo, Z., Jaggi, M., Javidi, T., Joshi, G., Kale, S., Karimireddy, S.P., Konecny, J., Koushanfar, F., Koyejo, O., Li, T., Liu, Y., Mohri, M., Nock, R., Pappas, G., Qi, Y., Reddi, S., Richtárik, P., Singhal, K., Smith, V., Suresh, A.T., Su, J., Sun, H., Talwalkar, A., Wang, H., Wu, L., Xu, S., Yang, Q., Yu, F.X., Yuan, M., Zaheer, M., Zhang, K. and Zhao, Z. (2021) 'Advances and open problems in federated learning', Foundations and Trends in Machine Learning, 14(1–2), pp. 1–210.

14. Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S., Stich, S.U. and Suresh, A.T. (2020) 'SCAFFOLD: Stochastic controlled averaging for federated learning', Proceedings of ICML, pp. 5132–5143.

15. Konečný, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T. and Bacon, D. (2016) 'Federated learning: Strategies for improving communication efficiency', arXiv preprint arXiv:1610.05492.

16. Li, T., Sahu, A.K., Talwalkar, A. and Smith, V. (2020) 'Federated optimization in heterogeneous networks', Proceedings of MLSys, pp. 429–450.

17. Lian, X., Zhang, C., Zhang, H., Hsieh, C.J., Zhang, W. and Liu, J. (2017) 'Can decentralized algorithms outperform centralized algorithms? A case study for decentralized parallel stochastic gradient descent', Advances in Neural Information Processing Systems, 30, pp. 5330–5340.

18. Liu, Y., Kang, Y., Xing, C., Chen, T. and Yang, Q. (2020) 'A secure federated transfer learning framework', IEEE Intelligent Systems, 35(4), pp. 70–82.

19. McMahan, B., Moore, E., Ramage, D., Hampson, S. and Arcas, B.A. (2017) 'Communication-efficient learning of deep networks from decentralized data', Proceedings of AISTATS, pp. 1273–1282.

20. Moustafa, N. and Slay, J. (2015) 'UNSW-NB15: A comprehensive data set for network intrusion detection systems', Military Communications and Information Systems Conference, pp. 1–6.

21. Nguyen, T.D., Marchal, S., Miettinen, M., Fereidooni, H. and Asokan, N. (2022) 'Deep federated learning for intrusion detection in IoT networks', IEEE Internet of Things Journal, 9(7), pp. 5625–5638.

22. Parisi, G.I., Kemker, R., Part, J.L., Kanan, C. and Wermter, S. (2019) 'Continual lifelong learning with neural networks: A review', Neural Networks, 113, pp. 54–71.

23. Pillutla, K., Kakade, S. and Harchaoui, Z. (2022) 'Robust aggregation for federated learning', IEEE Transactions on Signal Processing, 70, pp. 1142–1154.

24. Roesch, M. (1999) 'Snort: Lightweight intrusion detection for networks', Proceedings of LISA, pp. 229–238.

25. Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M. and Monfardini, G. (2009) 'The graph neural network model', IEEE Transactions on Neural Networks, 20(1), pp. 61–80.

26. Shi, W., Cao, J., Zhang, Q., Li, Y. and Xu, L. (2016) 'Edge computing: Vision and challenges', IEEE Internet of Things Journal, 3(5), pp. 637–646.

27. Shokri, R., Stronati, M., Song, C. and Shmatikov, V. (2017) 'Membership inference attacks against machine learning models', IEEE Symposium on Security and Privacy, pp. 3–18.

28. Sommer, R. and Paxson, V. (2010) 'Outside the closed world: On using machine learning for network intrusion detection', IEEE Symposium on Security and Privacy, pp. 305–316.

29. Truex, S., Liu, L., Chow, K.H., Gursoy, M.E. and Wei, W. (2019) 'Hybrid federated learning: Algorithms and implementation', IEEE International Conference on Big Data, pp. 3378–3387.

30. Yang, Q., Liu, Y., Chen, T. and Tong, Y. (2019) 'Federated machine learning: Concept and applications', ACM Transactions on Intelligent Systems and Technology, 10(2), pp. 1–19.

31. Yao, A.C. (1982) 'Protocols for secure computations', Proceedings of FOCS, pp. 160–164.

32. Yin, C., Zhu, Y., Fei, J. and He, X. (2017) 'A deep learning approach for intrusion detection using recurrent neural networks', IEEE Access, 5, pp. 21954–21961.

33. Zhang, Y., Kasahara, S., Shen, Y., Jiang, X. and Wan, J. (2020) 'Smart contract-based access control for the Internet of Things', IEEE Internet of Things Journal, 6(2), pp. 1594–1605.

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

35. Zissis, D. and Lekkas, D. (2012) 'Addressing cloud computing security issues', Future Generation Computer Systems, 28(3), pp. 583–592.

36. Bhagoji, A.N., Chakraborty, S., Mittal, P. and Calo, S., 2019. Analyzing federated learning through an adversarial lens. Proceedings of the 36th International Conference on Machine Learning (ICML), pp.634–643.

37. Cao, X., Fang, M., Liu, J. and Gong, N.Z., 2021. FLTrust: Byzantine-robust federated learning via trust bootstrapping. Proceedings of the Network and Distributed System Security Symposium (NDSS).