International Journal of Computer Science and Engineering Development
Volume 1, Issue 5 | November - December 2026 | www.ijcsed.com
ISSN: **3139-0862**

Privacy-Preserving Federated Intrusion Detection for Distributed IoT Environments: A Critical Review of Mechanisms, Threats and Open Research Gaps

Sujata Pattnaik[*1], Dr. Ridhika Sidhi[*2]
[*1] Research Scholar, Department of Computer Science, William Carey University, Meghalaya, India
[*2] Phd. Guide, Department of Computer Science, William Carey University, Meghalaya, India

Abstract
The rapid expansion of the Internet of Things (IoT) has produced large, heterogeneous and geographically dispersed networks whose devices differ in hardware capability, communication pattern and exposure to attack. Machine-learning-based intrusion detection systems (IDS) can recognise malicious traffic in such networks, but conventional designs depend on collecting raw traffic at a central server, which increases privacy exposure, communication load and the risk of a single high-value data repository. Federated learning (FL) offers an alternative in which IoT devices, gateways or edge nodes train a shared detection model while keeping raw traffic local. This paper presents a structured critical review of privacy-preserving federated intrusion detection for distributed IoT environments. It first characterises the IoT security setting and the main families of intrusion detection. It then examines representative federated IDS studies and recent surveys, analyses the privacy mechanisms used to protect model updates—differential privacy, secure aggregation, homomorphic encryption, encrypted communication and authentication—and describes the threats that target the federated learning process itself, including information leakage and model poisoning. The review shows that privacy, data heterogeneity, communication cost and robustness are usually studied in isolation, that privacy guarantees are rarely reported together with detection utility, and that many evaluations rely on simplified client partitions. Building on this analysis, the paper derives a set of research gaps and maps them to concrete research objectives and directions for an integrated, measurable and resource-aware federated IDS for IoT.

Index Terms—*Internet of Things, intrusion detection system, federated learning, differential privacy, secure aggregation, non-IID data, model poisoning, edge computing*

1. Introduction
The Internet of Things (IoT) connects sensors, smart appliances, industrial controllers, medical devices, vehicles and gateways through communication networks that support automation, remote monitoring and real-time decision-making. These benefits come with an expanded attack surface. IoT devices frequently run minimal operating systems, ship with weak default credentials, remain online for long periods and are rarely patched; large botnets such as Mirai have shown how quickly compromised IoT devices can be turned into instruments of distributed denial-of-service (DDoS) attack [1]. Unauthorised access, probing, abnormal communication and resource-exhaustion attacks can therefore threaten the availability, confidentiality and reliability of IoT services.

An intrusion detection system (IDS) monitors network traffic or host behaviour and raises alerts when activity indicates an attack. Learning-based IDS can model relationships in traffic data that are difficult to express as handcrafted signatures, and they have been widely studied for IoT networks [2], [3]. However, most learning-based IDS assume that traffic records from all monitored sites can be transferred to one server for training. In distributed IoT deployments this assumption is problematic: traffic reveals device behaviour, operational routines and sometimes user activity; moving it centrally increases bandwidth use on constrained links; and the resulting repository becomes an attractive target in its own right.

Federated learning (FL) addresses part of this problem by letting each participant train locally and share only model updates, which a server aggregates into a global model [4], [5]. FL is a natural fit for IoT, where data is already distributed across homes, factories, hospitals and edge sites [6]. Nevertheless, FL is not a complete privacy solution. Shared gradients and weights can leak information about the training data [7], [8], and malicious participants can manipulate the global model [9]. A federated IDS

ISSN: **3139-0862**

must therefore combine collaborative learning with explicit privacy and security mechanisms, while remaining practical under non-identically distributed (non-IID) client data and limited device resources.

Several surveys have examined FL for intrusion detection [10]–[14] and FL security more generally [15], [16]. This review complements them by organising the literature around four interacting requirements that a deployable IoT federated IDS must satisfy simultaneously: detection utility, privacy of local data and updates, robustness of the learning process, and resource efficiency under heterogeneity. The main contributions are as follows:

1) A structured characterisation of the distributed IoT security setting and of the intrusion detection approaches that a federated IDS builds upon.
2) A comparative analysis of representative federated IDS studies and of the privacy-preserving mechanisms used to protect model updates, including their protection scope, cost and suitability for IoT.
3) A threat model that separates attacks on the monitored IoT network from attacks on the federated learning process itself.
4) A synthesis of open research gaps, mapped to specific research objectives and future directions for integrated, measurable privacy-preserving federated intrusion detection.

Sections 2–3 present background and the review method; Sections 4–6 analyse federated IDS studies, privacy mechanisms and learning-process threats; Sections 7–9 derive research gaps, future directions and conclusions.

2. Background

2.1 Distributed and Heterogeneous IoT Environments

IoT deployments rarely behave as one uniform system. As illustrated in Fig. 1, a smart home produces low-volume, consumer-device traffic over Wi-Fi or Zigbee; an industrial site produces high-frequency telemetry over protocols such as Modbus or OPC UA with strict reliability requirements; a healthcare environment carries sensitive patient-related data with strong privacy constraints; and smart-city surveillance or connected vehicles generate high-bandwidth or highly mobile traffic. Edge and gateway devices aggregate these flows and typically have more computing capacity than end devices. Four forms of heterogeneity follow directly: traffic volume, computing resources, communication pattern and security exposure. Each one affects how a distributed IDS can be trained and evaluated.

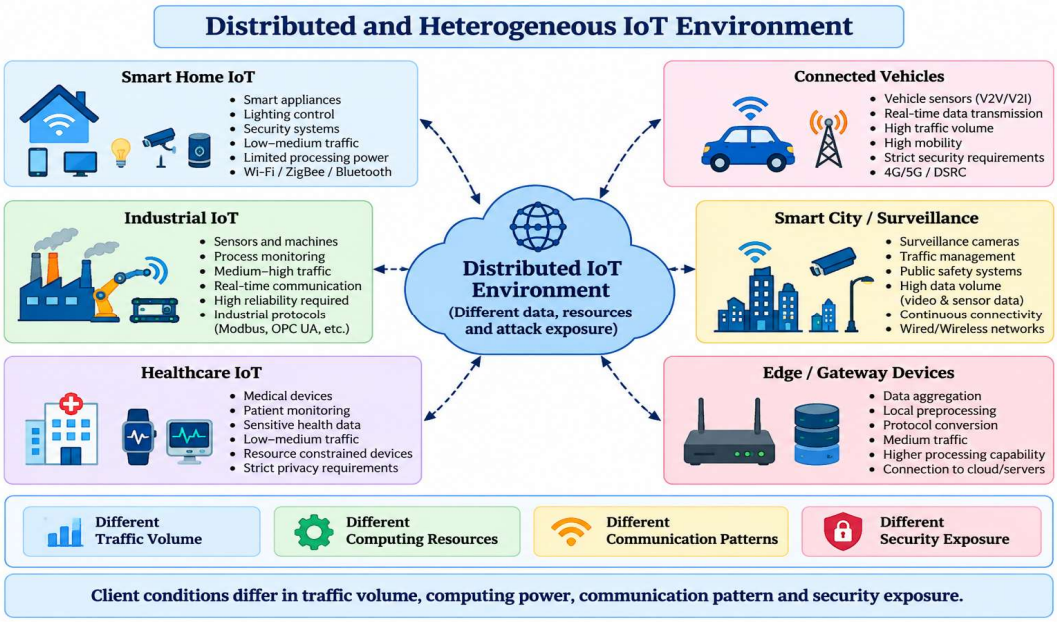


Fig. 1. Distributed and heterogeneous IoT environment. Smart-home, industrial, healthcare, smart-city and connected-vehicle clients differ in traffic volume, computing resources, communication patterns and security exposure; edge and gateway devices provide local aggregation and preprocessing.

2.2 Intrusion Detection Approaches for IoT

Intrusion detection approaches are commonly classified by where they observe activity and how they decide that activity is malicious (Fig. 2) [2]. Network-based IDS analyse packets and flows and are well suited to gateways and routers; host-based IDS inspect logs and behaviour on individual devices but impose overhead on constrained hardware; hybrid IDS combine both views. Signature-based

detection matches known attack patterns with low false-alarm rates but cannot recognise novel attacks, whereas anomaly-based and machine-learning detection learn normal or malicious behaviour from data and can generalise to new variants [3]. Lightweight learning detectors for IoT include autoencoder ensembles that run online on gateways [17] and deep autoencoders for botnet detection [18]. A long-standing caution is that learning-based IDS can perform well on benchmark data yet fail in operation because network traffic is diverse and the cost of false positives is high [19].

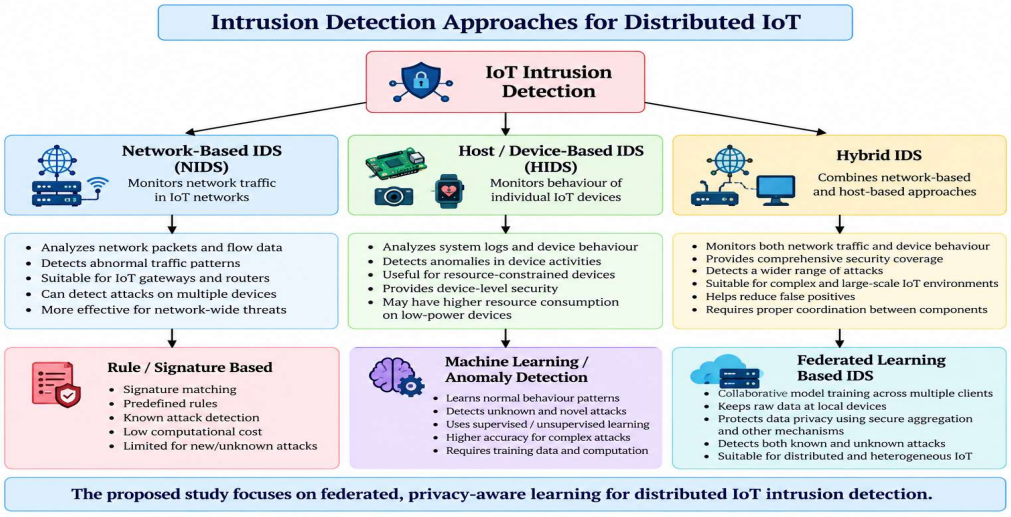


Fig. 2. Intrusion detection approaches for distributed IoT. Network-based, host-based and hybrid IDS differ in observation point; rule/signature-based, machine-learning/anomaly-based and federated-learning-based IDS differ in how detection knowledge is obtained.

2.3 Federated Learning and Federated Averaging

In FL, K clients hold local datasets D_k of size n_k, and the goal is to minimise the weighted global objective [4]

$$minw\ F(w) = \Sigma k = 1K\ \left(\frac{nk}{n}\right) Fk(w),\ \ Fk(w) = \left(\frac{1}{nk}\right) \Sigma(x,y) \in D_k\ \ell(w;\ x,y) \qquad (1)$$

where $n = \Sigma nk$ and ℓ is the per-sample loss. In each communication round t, the server sends the current model w^t to a subset S_t of clients; each client performs E local epochs of stochastic gradient descent to obtain w_k^{t+1}; and the server forms the new global model with federated averaging (FedAvg):

$$wt + 1 = \Sigma k \in S_t \left(\frac{nk}{nS_t}\right) wkt + 1 \qquad (2)$$

Only the parameters w_k^{t+1}—never the raw records in D_k—leave the client. Several local epochs between rounds reduce the number of communication rounds, which is attractive for bandwidth-limited IoT links, but they also allow local models to drift apart when client data distributions differ [5], [20].

2.4 Centralised Versus Federated Intrusion Detection

Table I contrasts centralised and federated IDS. The federated design changes where raw data resides and what is transmitted, but it introduces new design issues—non-IID data, update protection and robustness of aggregation—that are implicit or absent in the centralised case [21].

TABLE I COMPARISON OF CENTRALISED AND FEDERATED IDS FOR IOT

Aspect	Centralised IDS	Federated IDS
Data location	Single central repository	Remains on local clients (devices, gateways, edge nodes)
Raw-data transfer	Required for training	Not required; only model updates are exchanged
Privacy exposure	High: one repository holds traffic of all sites	Reduced raw-data exposure; updates may still leak information
Communication	Bulk traffic or feature transfer	Repeated model-update exchange over many rounds
Client heterogeneity	Hidden by pooling	Core design issue (non-IID data, unequal resources)
Learning-process attacks	Data poisoning at the source	Model/update poisoning by participating clients

3. Review Methodology

This work is a structured narrative review. Literature was identified from IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink and arXiv using combinations of the terms *federated learning*, *intrusion detection*, *anomaly detection*, *Internet of Things*, *differential privacy*,

secure aggregation, *non-IID* and *poisoning*. Three groups of sources were retained: (i) foundational work on FL, differential privacy and secure aggregation; (ii) peer-reviewed federated IDS or anomaly-detection studies evaluated on IoT or IoT-relevant traffic; and (iii) surveys published between 2019 and 2025 on FL-based intrusion detection and FL security. Each study was analysed along five themes derived from the requirements of a deployable system: privacy, data heterogeneity, communication cost, security of the learning process, and model suitability for constrained devices. The aim is analytical synthesis rather than an exhaustive count of publications.

4. Federated Intrusion Detection in IoT: State of the Art

4.1 Representative Federated IDS Studies

Early work established that federated training is feasible for IoT security. DÏoT trained device-type-specific anomaly-detection models through federated aggregation at security gateways, demonstrating that compromised IoT devices could be detected without pooling their traffic [22]. Rahman *et al.* compared centralised, on-device and federated intrusion detection and showed that federated training could approach centralised accuracy while avoiding raw-data sharing [21]. Subsequent studies applied recurrent and deep models to IoT attack detection in federated settings [23], examined zero-day botnet detection on IoT edge devices [24] and evaluated federated malware detection, including the effect of adversarial participants and robust aggregation [25]. Campos *et al.* evaluated federated IDS under different client distributions and aggregation functions and found that non-IID partitions can noticeably degrade detection, highlighting the gap between idealised and realistic federated settings [10]. Dataset efforts such as Edge-IIoTset were explicitly designed to support both centralised and federated evaluation [26]. Table II summarises these studies along the themes used in this review.

TABLE II LITERATURE MATRIX OF REPRESENTATIVE FEDERATED IDS STUDIES

Study	Focus / setting	Privacy & security treatment	Observation relevant to this review
Nguyen *et al.* [22]	Device-type anomaly detection at IoT gateways	Raw data local; no formal update protection	Shows federated detection of compromised devices is feasible
Rahman *et al.* [21]	Centralised vs on-device vs federated IDS	Raw data local	Federated accuracy can approach centralised; heterogeneity treated simply
Mothukuri *et al.* [23]	Recurrent models for IoT attack detection	Raw data local	Utility reported; privacy and communication cost not quantified together
Popoola *et al.* [24]	Zero-day botnet detection on IoT-edge devices	Raw data local	Highlights distributed knowledge sharing across edge sites
Rey *et al.* [25]	Federated malware detection in IoT	Considers adversarial clients and robust aggregation	Learning-process attacks are a practical concern
Campos *et al.* [10]	Evaluation under IID and non-IID partitions	Raw data local	Non-IID data and aggregation choice strongly affect outcomes

4.2 Survey Literature and Emerging Consensus

Recent surveys converge on several observations. Agrawal *et al.* and Campos *et al.* identify heterogeneity, communication and the lack of standard evaluation as central obstacles [10], [11]. Vyas *et al.* focus specifically on privacy-preserving FL for IoT intrusion detection and stress that FL alone does not guarantee privacy [12]. Li *et al.* review data security and privacy preservation for FL in edge-IoT [27], while Belenguer *et al.* and Zhang *et al.* catalogue the growing range of FL-IDS architectures, models and datasets [13], [14]. Khraisat *et al.* present a privacy-preserving FL strategy for IoT intrusion detection that integrates protection into the training workflow [28]. Across these works the common message is that the individual building blocks are known, but they are rarely combined and measured together within one reproducible experimental design.

5. Privacy-Preserving Mechanisms for Federated IDS

5.1 Why Model Updates Need Protection

Sharing gradients instead of data reduces but does not eliminate disclosure. Collaborative-learning updates can reveal properties of participants' training data that are unrelated to the learning task [7], and under some conditions training samples can be reconstructed from shared gradients [8]. In an IoT IDS, such leakage could expose device inventories, operating schedules or the attack history of a site. A layered defence is therefore required, as summarised in Fig. 3: raw data is retained locally, updates are clipped and perturbed, aggregation hides individual contributions, communication is encrypted, and only authenticated clients participate.

ISSN: **3139-0862**

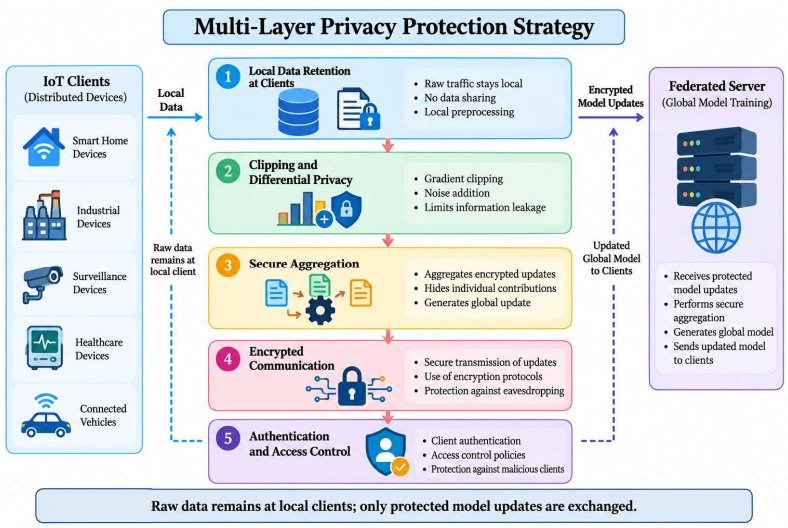


Fig. 3. Multi-layer privacy protection strategy for federated IDS: (1) local data retention, (2) clipping and differential privacy, (3) secure aggregation, (4) encrypted communication and (5) authentication and access control. Raw data remains at clients; only protected model updates are exchanged.

5.2 Differential Privacy

Differential privacy (DP) bounds how much any single record can influence a computation's output [29]. A randomised mechanism M is (ε, δ)-differentially private if, for all neighbouring datasets D and D' and all output sets S,

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \cdot \Pr[M(D') \in S] + \delta \quad (3)$$

Smaller ε and δ indicate stronger privacy. In FL, DP is typically applied by clipping each update to norm C and adding Gaussian noise before release [30]:

$$\tilde{\Delta}_k = \Delta_k / \max(1, \|\Delta_k\|_2 / C) + N(0, \sigma^2 C^2 I) \quad (4)$$

where $\Delta_k = w_k^{t+1} - w^t$ and σ is the noise multiplier. Client-level DP protects the participation of an entire client, whereas record-level DP protects individual flows. Wei *et al.* analysed the convergence of DP-based FL and showed an explicit trade-off between privacy level, number of participating clients and model performance [31]. For intrusion detection this trade-off is critical: attack classes are often rare, and noise can disproportionately reduce recall on minority attacks.

5.3 Secure Aggregation and Homomorphic Encryption

Secure aggregation allows the server to learn only the sum of client updates. In the protocol of Bonawitz *et al.*, clients add pairwise masks that cancel in the sum, with secret sharing used to tolerate dropouts [32]. Additively homomorphic encryption achieves a similar goal by letting the server add encrypted updates without decryption [33]. Both protect individual contributions from an honest-but-curious server but do not protect the aggregate itself, which is why they are often combined with DP. Their costs—extra communication rounds for key agreement, ciphertext expansion and computation—matter on constrained IoT clients.

5.4 Encrypted Communication, Authentication and Access Control

Transport-layer encryption protects updates from eavesdropping and tampering in transit, while authentication and access control ensure that only enrolled clients participate. These mechanisms do not provide formal privacy guarantees on their own, but they reduce the opportunity for impersonation and Sybil participation, which are prerequisites for many poisoning attacks [15]. Table III compares the mechanisms discussed in this section.

TABLE III COMPARISON OF PRIVACY AND SECURITY MECHANISMS FOR FEDERATED IDS

Mechanism	Protects against	Main cost	IoT suitability / caveat
Local data retention (FL)	Direct exposure of raw traffic	Repeated update exchange	Baseline; insufficient alone [7], [8]
Differential privacy	Inference of records or clients from updates or the model	Utility loss, especially on rare attack classes	Lightweight to compute; requires privacy accounting [30]
Secure aggregation	Inspection of individual updates by the server	Extra key-agreement rounds and messages	Must tolerate dropout of unreliable devices [32]
Homomorphic encryption	Server access to plaintext updates	Ciphertext expansion, heavy computation	Often too costly for end devices; feasible at gateways [33]

ISSN: **3139-0862**

Encrypted channels	Eavesdropping and tampering in transit	Handshake and encryption overhead	Standard practice; no protection at endpoints
Authentication & access control	Unauthorised or Sybil participants	Credential management	Precondition for robust aggregation

6. Threats to the Federated Learning Process

A federated IDS faces two distinct classes of threat (Fig. 4). The first is the set of attacks on the monitored IoT network—scanning, DoS and DDoS, exploitation of vulnerable services and unauthorised access—which the IDS is designed to detect. The second targets the federated system itself: information leakage from updates, communication attacks such as eavesdropping and man-in-the-middle manipulation, compromised clients that submit malicious updates, and insiders who abuse legitimate access [15], [16].

6.1 Model Poisoning and Backdoors

Because the server cannot inspect local data, a compromised client can submit updates that degrade the global model or implant a backdoor. Bagdasaryan *et al.* showed that model replacement can insert a targeted backdoor while preserving main-task accuracy [9], and Fang *et al.* demonstrated local model-poisoning attacks that remain effective against several Byzantine-robust aggregation rules [34]. In an IDS, a backdoor could cause a specific attack type to be classified as benign—a particularly damaging outcome. Malware-detection studies in IoT confirm that such adversarial participants are a practical concern [25].

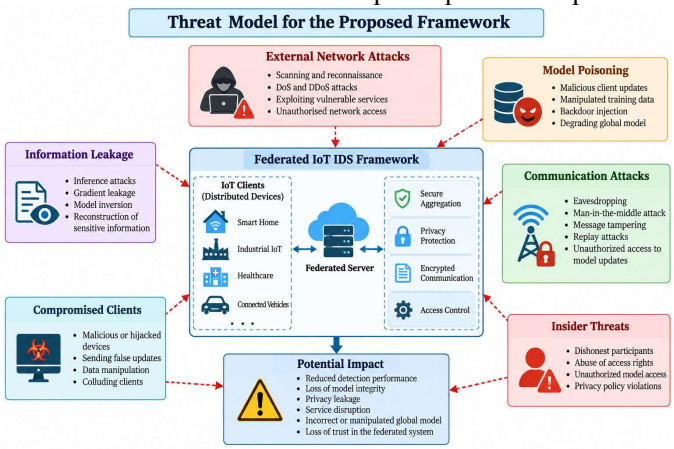


Fig. 4. Threat model for a federated IDS. External network attacks are the detection target; information leakage, communication attacks, compromised clients, model poisoning and insider threats target the federated learning infrastructure.

6.2 Robust Aggregation

Robust aggregation replaces the weighted mean in (2) with statistics that limit the influence of outliers. Coordinate-wise median and trimmed mean are representative examples [35]; for a trimming parameter β, the trimmed mean removes the β fraction of largest and smallest values in each coordinate before averaging:

$$w^{t+1}_j = \mathrm{TrMean}_\beta\{ w^{t+1}_{k,j} : k \in S_t \}, \quad j = 1, \ldots, d \qquad (5)$$

Robust rules assume that malicious clients are a minority and that benign updates are similar. Both assumptions weaken under non-IID IoT data, where legitimate clients may also look like outliers, and secure aggregation conflicts with robust rules that require access to individual updates. Reconciling privacy-preserving aggregation with robustness therefore remains an open problem [16].

7. Open Challenges and Research Gaps

The preceding analysis shows that the core problems of a federated IDS are not independent: privacy of local traffic data, data heterogeneity, communication overhead and security of the learning process jointly determine whether a privacy-preserving federated IDS is useful in practice. The gaps in each dimension are discussed below.

7.1 Non-IID Data and Client Heterogeneity

IoT clients naturally hold different attack mixes and volumes. FedAvg can lose substantial accuracy when client label distributions are skewed [20], and remedies such as proximal regularisation (FedProx) [36], control variates (SCAFFOLD) [37] have been proposed. Their behaviour on intrusion data, where some attack classes may exist at only a few clients, is less well documented [10]. Averages over all clients can also hide poor detection at an individual site, so per-client reporting is needed.

7.2 Communication and Resource Constraints

Repeated exchange of model updates can dominate the cost of FL on IoT links. Structured and sketched updates, quantisation and sparsification reduce upload size [38], [39], and smaller detection models reduce both computation and update size. However, many FL-IDS papers report accuracy but not the number of rounds, bytes exchanged, training time or model size, which prevents fair comparison of resource efficiency [11], [13].

7.3 Measuring the Privacy–Utility Trade-off

Studies often state that FL is privacy-preserving without quantifying the protection actually provided. Where DP is used, the privacy budget, clipping bound and noise multiplier are not always reported together with detection metrics [12], [31]. For IDS in particular, the effect of noise on minority-class recall and on false-positive rate deserves explicit reporting rather than aggregate accuracy alone.

7.4 Evaluation Practice and Datasets

Learning-based IDS are vulnerable to over-optimistic evaluation [19]. Modern IoT datasets—CICIoT2023 [40], TON_IoT [41], IoT-23 [42] and Edge-IIoTset [26]—offer more realistic devices and attacks than legacy benchmarks, yet federated evaluations frequently split them uniformly at random, which does not reflect how traffic is distributed across real sites. Threat models for poisoning or leakage are also often implicit, making robustness claims hard to interpret.

7.5 Synthesis: From Gaps to Research Objectives

Fig. 5 maps the identified gaps to a set of research objectives for an integrated framework: analysing IoT security characteristics, designing a federated IDS architecture, integrating privacy mechanisms, and developing a lightweight model evaluated under realistic distributed conditions. Table IV consolidates the gaps and states how each can be addressed.



Fig. 5. Mapping from existing work and identified research gaps to the research aim and objectives of an integrated privacy-preserving federated IDS for IoT.

TABLE IV RESEARCH GAPS IDENTIFIED AND CORRESPONDING RESEARCH RESPONSE

Theme	Common observation in literature	Gap to be addressed
Privacy	FL reduces raw-data sharing [12], [27]	Measure privacy (ε, δ, clipping, noise) together with detection utility
Non-IID data	Client differences affect learning [10], [20]	Use controlled, documented non-IID partitions and per-client reporting
Communication	Model exchange has a cost [38]	Report rounds, update size, bytes exchanged and training time
Security	Poisoning and leakage are possible [9], [16]	Define an explicit threat model and test selected attacks
Models	Lightweight learning is useful on edge devices [17]	Resource-aware model selection with reported model size
Evaluation	Centralised, FL and privacy-preserving FL rarely compared fairly	Compare all three under identical data, splits and metrics

8. Future Research Directions

The gaps above suggest five priority directions for deployable privacy-preserving federated IDS:

- **Integrated, multidimensional evaluation.** Report detection metrics (precision, recall, F1-score, false-positive rate, AUC) together with privacy parameters, communication cost and training time, so that trade-offs become visible rather than assumed.
- **Heterogeneity-aware federated optimisation for IDS.** Evaluate FedProx-, SCAFFOLD- and

personalisation-style methods on intrusion data with rare, client-specific attack classes, including partial participation and dropout.

- **Compatible privacy and robustness.** Develop aggregation schemes that retain secure-aggregation guarantees while limiting the influence of poisoned updates, for example by validation on protected statistics or by clustering at trusted gateways.
- **Resource-aware models and hierarchical FL.** Use lightweight local detectors on end devices with aggregation at gateways or edge servers to reduce wide-area communication.
- **Reproducibility.** Publish dataset versions, partitioning procedures, random seeds, privacy accounting and code to allow independent verification.

9. Conclusion

This review examined privacy-preserving federated intrusion detection for distributed IoT environments. Federated learning removes the need to centralise raw traffic and is therefore well aligned with the distributed nature of IoT, but it does not by itself guarantee privacy, robustness or efficiency. The literature provides mature building blocks—federated averaging, differential privacy, secure aggregation, homomorphic encryption and robust aggregation—yet these are usually evaluated in isolation, privacy is often asserted rather than measured, and client heterogeneity is frequently simplified. The research gaps identified here point toward an integrated framework in which local preprocessing, lightweight local detection, layered update protection and federated aggregation are designed and evaluated together under explicit threat models and realistic non-IID conditions.

References

[1] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and other botnets," *Computer*, vol. 50, no. 7, pp. 80–84, 2017, doi: 10.1109/MC.2017.201.

[2] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: Techniques, datasets and challenges," *Cybersecurity*, vol. 2, Art. no. 20, 2019, doi: 10.1186/s42400-019-0038-7.

[3] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, "Network intrusion detection for IoT security based on learning techniques," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2671–2701, 2019, doi: 10.1109/COMST.2019.2896380.

[4] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS)*, Fort Lauderdale, FL, USA, 2017, pp. 1273–1282, doi: 10.48550/arXiv.1602.05629.

[5] P. Kairouz *et al.*, "Advances and open problems in federated learning," *Found. Trends Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, 2021, doi: 10.1561/2200000083.

[6] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for Internet of Things: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1622–1658, 2021, doi: 10.1109/COMST.2021.3075439.

[7] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in *Proc. IEEE Symp. Secur. Privacy (SP)*, San Francisco, CA, USA, 2019, pp. 691–706, doi: 10.1109/SP.2019.00029.

[8] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, Vancouver, BC, Canada, 2019, doi: 10.48550/arXiv.1906.08935.

[9] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," in *Proc. 23rd Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2020, pp. 2938–2948, doi: 10.48550/arXiv.1807.00459.

[10] E. M. Campos *et al.*, "Evaluating federated learning for intrusion detection in Internet of Things: Review and challenges," *Comput. Netw.*, vol. 203, Art. no. 108661, 2022, doi: 10.1016/j.comnet.2021.108661.

[11] S. Agrawal *et al.*, "Federated learning for intrusion detection system: Concepts, challenges and future directions," *Comput. Commun.*, vol. 195, pp. 346–361, 2022, doi: 10.1016/j.comcom.2022.09.012.

[12] A. Vyas, P.-C. Lin, R.-H. Hwang, and M. Tripathi, "Privacy-preserving federated learning for intrusion detection in IoT environments: A survey," *IEEE Access*, vol. 12, pp. 127018–127050, 2024, doi: 10.1109/ACCESS.2024.3454211.

[13] A. Belenguer, J. A. Pascual, and J. Navaridas, "A review of federated learning applications in intrusion detection systems," *Comput. Netw.*, vol. 258, Art. no. 111023, 2025, doi: 10.1016/j.comnet.2024.111023.

[14] H. Zhang, J. Ye, W. Huang, X. Liu, and J. Gu, "Survey of federated learning in intrusion detection," *J. Parallel Distrib. Comput.*, vol. 195, Art. no. 104976, 2025, doi: 10.1016/j.jpdc.2024.104976.

[15] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Gener. Comput. Syst.*, vol. 115, pp. 619–640, 2021, doi: 10.1016/j.future.2020.10.007.

[16] L. Lyu *et al.*, "Privacy and robustness in federated learning: Attacks and defenses," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 7, pp. 8726–8746, 2024, doi: 10.1109/TNNLS.2022.3216981.

[17] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: An ensemble of autoencoders for online network intrusion detection," in *Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, San Diego, CA, USA, 2018, doi: 10.14722/ndss.2018.23204.

[18] Y. Meidan *et al.*, "N-BaIoT—Network-based detection of IoT botnet attacks using deep autoencoders," *IEEE Pervasive Comput.*, vol. 17, no. 3, pp. 12–22, 2018, doi: 10.1109/MPRV.2018.03367731.

[19] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Proc. IEEE Symp. Secur. Privacy*, Oakland, CA, USA, 2010, pp. 305–316, doi: 10.1109/SP.2010.25.

[20] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-IID data," 2018, *arXiv:1806.00582*, doi: 10.48550/arXiv.1806.00582.

[21] S. A. Rahman, H. Tout, C. Talhi, and A. Mourad, "Internet of Things intrusion detection: Centralized, on-device, or federated learning?," *IEEE Netw.*, vol. 34, no. 6, pp. 310–317, 2020, doi: 10.1109/MNET.011.2000286.

[22] T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan, and A.-R. Sadeghi, "DÏoT: A federated self-learning anomaly detection system for IoT," in *Proc. IEEE 39th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Dallas, TX, USA, 2019, pp. 756–767, doi: 10.1109/ICDCS.2019.00080.

[23] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriyeh, A. Dehghantanha, and G. Srivastava, "Federated-learning-based anomaly detection for IoT security attacks," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2545–2554, 2022, doi: 10.1109/JIOT.2021.3077803.

[24] S. I. Popoola, R. Ande, B. Adebisi, G. Gui, M. Hammoudeh, and O. Jogunola, "Federated deep learning for zero-day botnet attack detection in IoT-edge devices," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3930–3944, 2022, doi: 10.1109/JIOT.2021.3100755.

[25] V. Rey, P. M. Sánchez Sánchez, A. Huertas Celdrán, and G. Bovet, "Federated learning for malware detection in IoT devices," *Comput. Netw.*, vol. 204, Art. no. 108693, 2022, doi: 10.1016/j.comnet.2021.108693.

[26] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022, doi: 10.1109/ACCESS.2022.3165809.

[27] H. Li, L. Ge, and L. Tian, "Survey: Federated learning data security and privacy-preserving in edge-Internet of Things," *Artif. Intell. Rev.*, vol. 57, Art. no. 130, 2024, doi: 10.1007/s10462-024-10774-7.

[28] A. Khraisat *et al.*, "Federated learning for intrusion detection in IoT environments: A privacy-preserving strategy," *Discover Internet Things*, vol. 5, Art. no. 72, 2025, doi: 10.1007/s43926-025-00169-7.

[29] C. Dwork, "Differential privacy," in *Proc. 33rd Int. Colloq. Automata, Lang. Program. (ICALP)*, Venice, Italy, 2006, pp. 1–12, doi: 10.1007/11787006_1.

[30] M. Abadi *et al.*, "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, Vienna, Austria, 2016, pp. 308–318, doi: 10.1145/2976749.2978318.

[31] K. Wei *et al.*, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3454–3469, 2020, doi: 10.1109/TIFS.2020.2988575.

[32] K. Bonawitz *et al.*, "Practical secure aggregation for privacy-preserving machine learning," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, Dallas, TX, USA, 2017, pp. 1175–1191, doi: 10.1145/3133956.3133982.

[33] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1333–1345, 2018, doi: 10.1109/TIFS.2017.2787987.

[34] M. Fang, X. Cao, J. Jia, and N. Gong, "Local model poisoning attacks to Byzantine-robust federated learning," in *Proc. 29th USENIX Secur. Symp.*, 2020, pp. 1605–1622, doi: 10.48550/arXiv.1911.11815.

[35] D. Yin, Y. Chen, K. Ramchandran, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, Stockholm, Sweden, 2018, pp. 5650–5659, doi: 10.48550/arXiv.1803.01498.

[36] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst. (MLSys)*, vol. 2, 2020, pp. 429–450, doi: 10.48550/arXiv.1812.06127.

[37] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in *Proc. 37th Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 5132–5143, doi: 10.48550/arXiv.1910.06378.

[38] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016, *arXiv:1610.05492*, doi: 10.48550/arXiv.1610.05492.

[39] F. Sattler, S. Wiedemann, K.-R. Müller, and W. Samek, "Robust and communication-efficient federated learning from non-i.i.d. data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3400–3413, 2020, doi: 10.1109/TNNLS.2019.2944481.

ISSN: **3139-0862**

[40] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," *Sensors*, vol. 23, no. 13, Art. no. 5941, 2023, doi: 10.3390/s23135941.

[41] N. Moustafa, "A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets," *Sustain. Cities Soc.*, vol. 72, Art. no. 102994, 2021, doi: 10.1016/j.scs.2021.102994.

[42] S. Garcia, A. Parmisano, and M. J. Erquiaga, "IoT-23: A labeled dataset with malicious and benign IoT network traffic," Zenodo, 2020, doi: 10.5281/zenodo.4743746.