Computer Science

E-ISSN No : 2581-6195 | Volume : 9 | Special Issue | Aug 2026

PRIVACY-PRESERVING INTRUSION DETECTION IN IOT USING FEDERATED LEARNING: A REVIEW

Mehul Kapoor[1], Dr. Manish Pandey[2]

ABSTRACT

The Internet of Things (IoT) is rapidly growing and consequently adding to the attack surface of connected infrastructures, making intrusion detection an important requirement. Traditional IDSs are designed to gather network traffic in a central location and involve a lot of privacy issues as well as clash with data protection laws. Federated learning (FL) has been proposed as a viable paradigm to train a global detection model across local IoT devices while retaining the raw data at the local server. In this review, the recent developments of privacy-preserving intrusion detection in IoT based on federated learning are presented. We outline the Federated training pipeline, aggregation of local updates, and mechanisms to make it resist inference and poisoning attacks, namely differential privacy, secure aggregation and homomorphic encryption. We also discuss representative studies, provide summary of some widely used datasets and an analysis of accuracy–privacy trade-off. Finally we state some open problems concerning heterogeneity, communication cost and robustness and we point out interesting research lines.

KEYWORDS: Federated Learning, Intrusion Detection, Internet Of Things, Privacy Preservation, Differential Privacy, Secure Aggregation, Network Security

[1] Research Scholar, Department of Computer Science and Engineering, Phonics University, Roorkee, India

[2] Department of Computer Science and Engineering, Phonics University, Roorkee, India

How to Cite:
Mehul Kapoor & Dr. Manish Pandey (2026), Privacy-Preserving Intrusion Detection in IoT Using Federated Learning: A Review, International Educational Journal of Science & Engineering (IEJSE),Vol: 9, Special Issue, Aug 2026 101-105

I. INTRODUCTION

Today, billions of resource limited devices are connected in smart homes, Industry, Healthcare and Transportation via the Internet of Things (IoT). This scale, combined with weak security on devices, makes it easy for large-scale attacks like botnets, denial of service, and reconnaissance to target an IoT deployment. Intrusion detection systems (IDS) are then key to IoT defense as they detect malicious actions before they spread across a network [1],[2].

The traditional IDS designs also rely on the premise that traffic from all nodes can be collected at a central server to be trained and inferred. This is a questionable assumption in an IoT context. When the raw telemetry is moved off the device, it reveals sensitive behavioral information, the data communication cost is high and often contravenes the data-protection policies like GDPR [9]. The limitations have encouraged a change to collaborative learning with data left in place of generation.

To address this need, federated learning (FL) has been proposed to train a shared model from many devices without sharing their local data [1, 8]. All of the devices process their own traffic and only upload update information to an aggregation device. This means that FL is naturally complementary with privacy requirements of intrusion detection in IoT systems [3] and [4]. The exchanged updates, however, can also leak information, and often FL is used together with privacy-preserving mechanisms.

This paper provides a review of the existing research on privacy-preserving

Research Paper

intrusion detection in IoT using federated learning. The rest of the paper is organized as follows. In Section II, background on IoT security and FL are provided. The federated IDS pipeline and its basic equations are described in Section III. In Section IV privacy preserving techniques are discussed. Data and representative studies are compared in the content of Section V in the form of Tables I-III and Figs. 1–2. Section VI presents open challenges and future directions and Section VII concludes.

II. BACKGROUND

A. IoT Security and Intrusion Detection

IoT intrusion detection is a method used to detect if network flows or the telemetry of devices aren't acting as they would under normal circumstances. Signature-based detectors recognize known patterns but do not work well with new attacks, while anomaly-based and learning-based detectors are more general and outperform the previous studies [5] and [6]. Deep models like convolutional and recurrent networks have shown high accuracy, but generally they are built on data specific to a central entity, something that privacy constraints prohibit [13].

B. Federated Learning

Federated learning coordinates a set of clients and a central server over multiple communication rounds [8]. In each round the server broadcasts the current global model, clients perform local training, and the server aggregates the returned updates. The canonical aggregation rule, federated averaging, weights each client update by its local sample count, as given in (1).

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}{}^{k} \quad (1)$$

Here w denotes the global model at round t, K is the number of participating clients, and n_k is the number of samples on client k out of n total. Each client minimizes a local objective built from its private dataset D_k, expressed in (2), where ℓ is the per-sample loss.

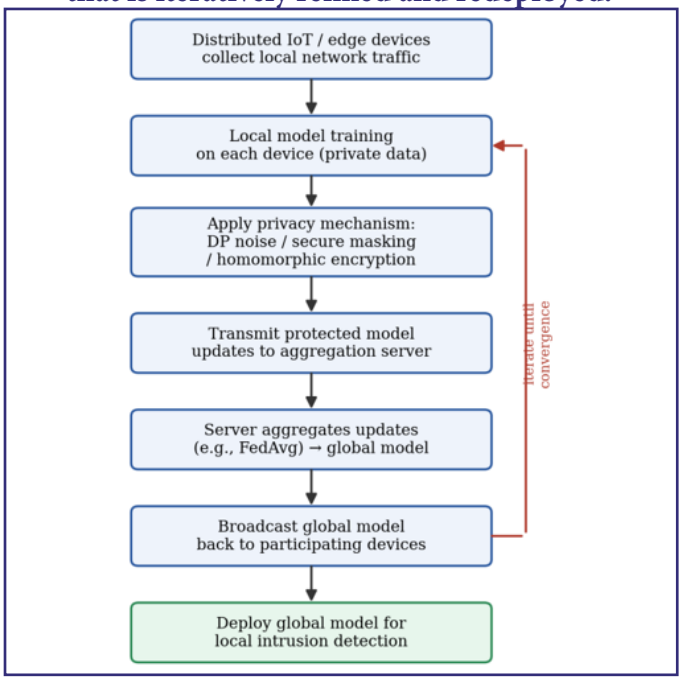
$$F_k(w) = \frac{1}{n_k} \sum_{i \in D_k} \ell(w;\ x_i, y_i) \quad (2)$$

Federated averaging reduces communication compared with sharing gradients every step and tolerates the statistical heterogeneity typical of IoT fleets, though non-independent and identically distributed (non-IID) data can still slow convergence [2], [8].

III. FEDERATED INTRUSION DETECTION PIPELINE

A federated IDS instantiates the FL workflow with detection models trained on local traffic. Fig. 1 summarizes the pipeline. Distributed IoT or edge nodes first collect local traffic; each node then trains a detection model on its private data. Before transmission, a privacy mechanism perturbs or encrypts the update; the server aggregates the protected updates into an improved global model and broadcasts it back. The loop repeats until convergence, after which the global model is deployed for local detection [5], [7], [21].

Fig. 1: Privacy-preserving federated learning pipeline for IoT intrusion detection. Protected local updates are aggregated into a global model that is iteratively refined and redeployed.



As illustrated in Fig. 1, the only information leaving a device is a protected model update, never raw traffic. This structure keeps the detector current as new attack behavior appears on individual nodes, while confining sensitive data to its origin. Detection quality is reported using standard classification metrics; the

F1 score in (6) balances precision P and recall R and is preferred when attack classes are imbalanced.

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{6}$$

IV. PRIVACY-PRESERVING TECHNIQUES

Although FL avoids raw-data sharing, model updates can still be inverted to recover sensitive information [3], [9]. Three complementary mechanisms are used to close this gap, summarized in Table II.

A. Differential Privacy

Differential privacy (DP) bounds the influence of any single record by adding calibrated noise. A mechanism M satisfies (ε, δ)-DP if, for neighboring datasets D and D', the condition in (3) holds. Smaller ε yields stronger privacy but greater accuracy loss [10], [22].

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

In practice, Gaussian noise scaled to the update sensitivity S is added to each client update before transmission, as in (4). Wei et al. [10] derived convergence bounds that quantify how this noise degrades accuracy, exposing the fundamental privacy–utility trade-off.

$$\tilde{w}_k = w_k + N(0, \sigma^2 S^2) \tag{4}$$

B. Secure Aggregation

Secure aggregation lets the server compute the sum of client updates without seeing any individual contribution. Clients add pairwise masks that cancel once all updates are summed, as expressed in (5), so the server recovers only the aggregate [3], [4]. The main cost is additional communication and the need to recover from client dropout.

$$\sum_k (w_k + s_k) = \sum_k w_k, \ \sum_k s_k = 0 \tag{5}$$

C. Homomorphic Encryption and Ledgers

Homomorphic encryption enables the server to aggregate over ciphertexts, offering strong confidentiality at high computational cost that strains constrained devices [3]. Blockchain and distributed-ledger approaches add tamper-evident, verifiable exchange of updates and have been applied to federated IDS in smart transportation [17]. These mechanisms are often combined; for example, DP can be layered on top of secure aggregation.

V. DATASETS AND COMPARATIVE ANALYSIS

Reproducible evaluation depends on representative datasets. Table I lists benchmarks commonly used for federated IoT intrusion detection, spanning botnet, industrial, and general network traffic. TON_IoT and Edge-IIoTset are of particular interest because they were designed with heterogeneous, edge-oriented deployments in mind [11], [12].

Table I: Datasets Used In Federated Iot Intrusion Detection

Dataset	Year / Ref.	Salient characteristics
N-BaIoT	2020 [21]	Botnet traffic from nine commercial IoT devices; benign and Mirai/Gafgyt attacks.
TON_IoT	2021 [11], [12]	Heterogeneous telemetry, network and OS logs; nine attack families for edge AI.
Edge-IIoTset	2022 [7]	Industrial IoT flows with 14 attack types; designed for federated evaluation.
CIC-IDS2017	2020 use [18]	Realistic benign/attack flows widely reused as an IoT-IDS benchmark.
BoT-IoT	2020 use [6]	Large-scale IoT botnet dataset with DoS, DDoS, reconnaissance and theft.

Table I compares five widely used datasets by year, source, and salient characteristics relevant to federated evaluation.

Table II contrasts the privacy-preserving techniques of Section IV by protection principle and dominant trade-off. As the table shows, no single mechanism is universally optimal: differential privacy is lightweight but degrades accuracy, secure aggregation preserves accuracy but adds communication overhead, and homomorphic encryption offers the strongest confidentiality at the highest computational cost.

Table II: Comparison of Privacy-Preserving Mechanisms

Technique	Protection principle	Main trade-off
Differential privacy	Calibrated noise bounds the influence of any record (ε, δ).	Accuracy loss grows as ε shrinks.
Secure aggregation	Additive masks hide individual updates; masks cancel on sum.	Communication and dropout-recovery overhead.
Homomorphic encryption	Server aggregates over ciphertexts without decryption.	High computational cost on devices.
Blockchain / ledger	Tamper-evident, verifiable model exchange.	Latency and storage overhead.

Table II summarizes each mechanism's protection principle and principal trade-off, as discussed in Section IV.

Table III consolidates representative studies, listing the detection model, the privacy setting, and the reported outcome. The surveyed work confirms that federated detectors can approach centralized accuracy while retaining data locality [5], [6], and that adding differential privacy introduces a measurable but often acceptable accuracy penalty [10], [19].

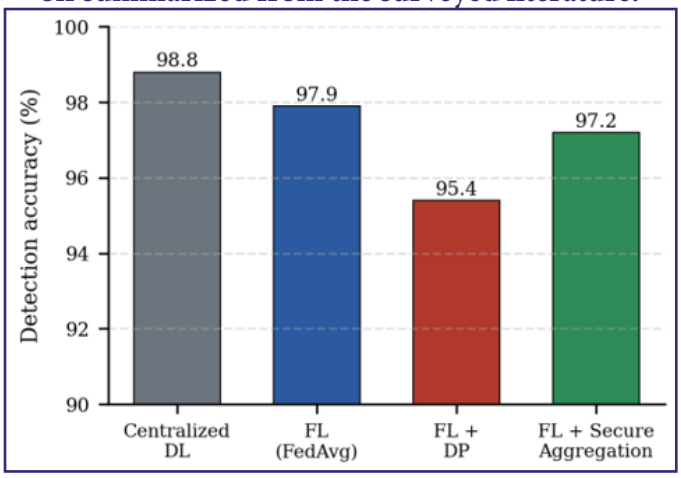
Table III: Representative Federated Iot-Ids Studies

Study	Model	Privacy	Reported outcome
Mothukuri [5]	GRU ensemble	FL only	Federated anomaly detection near centralized accuracy.
Ferrag [6]	Deep NN	FL only	FL evaluated across three IoT datasets.
Ruzafa [19]	MLP	FL + DP	Privacy-utility trade-off quantified for IIoT.
Friha [7]	CNN/RNN	FL only	FELIDS for agricultural IoT edge nodes.
Wei [10]	Generic	FL + DP	Convergence bound under DP noise.

Table III maps five representative studies to their model, privacy setting, and headline result.

Fig. 2 visualizes the accuracy–privacy trade-off using indicative results synthesized from the surveyed studies. A centralized deep detector attains the highest accuracy but forfeits privacy. Plain federated averaging retains most of that accuracy while keeping data local. Adding differential privacy produces the largest drop, reflecting the noise in (4), whereas secure aggregation preserves accuracy at the cost of communication overhead [4], [10], [19].

Fig. 2: Indicative detection accuracy across four settings, illustrating the accuracy–privacy trade-off summarized from the surveyed literature.



As Fig. 2 indicates, the practical design question is not whether to protect privacy but how much accuracy a deployment can afford to trade for a target privacy guarantee. This choice depends on the threat model, the sensitivity of the traffic, and the resource budget of the devices [9], [13].

VI. CHALLENGES AND FUTURE DIRECTIONS

There are a number of challenges yet to be solved. First, IoT data exhibit high non-IID and unbalanced nature across devices, thereby creating a slowdown in federated convergence and leading to a bias of the global detector [2, 16]. Second, resource and communication constraints, which will make exchanging models repeatedly expensive, drive model compression and selective client participation [7, 23]. Third, FL is susceptible to poisoning attacks where malicious clients alter the global model and to backdoor attacks, where updates from malicious clients are manipulated and added [15] [17]. Fourth, the privacy-utility trade-off should be adapted depending on the deployment: If the noise is too high, the detector may not detect anything [10], [19].

Promising directions are personalized and semi-supervised federated detection that adapts to each device while leveraging limited labels [16], lightweight cryptography for limited hardware [17] and verifiability with limited latency using ledgers [20]. For future work, it will be important to use benchmarks that are standardized and heterogeneous like TON_IoT and Edge-IIoTset in order to fairly

compare the results of different research efforts [11, 12].

VII. CONCLUSION

Federated learning provides a principled approach for intrusion detection in IoT, while maintaining data locality and privacy. Recent systems increase the accuracy of distributed devices by training shared detectors, while also adding differential privacy, secure aggregation, or homomorphic encryption to reinforce the accuracy. Achieving this goal at scale, however, still requires work on the issues of heterogeneity, communication efficiency and robustness against malicious clients. How well privacy preserving federated intrusion detection can be achieved to address these issues and with adherence to standardized benchmarks will be the difference between the success and failure for the next generation of connected systems.

REFERENCES

1. P. Kairouz et al., "Advances and open problems in federated learning," Found. Trends Mach. Learn., vol. 14, no. 1–2, pp. 1–210, 2021.
2. 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.
3. 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.
4. E. M. Campos et al., "Evaluating federated learning for intrusion detection in Internet of Things: Review and challenges," Comput. Netw., vol. 203, art. 108661, 2022.
5. 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.
6. M. A. Ferrag, O. Friha, L. Maglaras, H. Janicke, and L. Shu, "Federated deep learning for cyber security in the Internet of Things: Concepts, applications, and experimental analysis," IEEE Access, vol. 9, pp. 138509–138542, 2021.
7. O. Friha, M. A. Ferrag, L. Shu, L. Maglaras, K.-K. R. Choo, and M. Nafaa, "FELIDS: Federated learning-based intrusion detection system for agricultural Internet of Things," J. Parallel Distrib. Comput., vol. 165, pp. 17–31, 2022.
8. T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," IEEE Signal Process. Mag., vol. 37, no. 3, pp. 50–60, 2020.
9. N. Truong, K. Sun, S. Wang, F. Guitton, and Y. Guo, "Privacy preservation in federated learning: An insightful survey from the GDPR perspective," Comput. Secur., vol. 110, art. 102402, 2021.
10. K. Wei et al., "Federated learning with differential privacy: Algorithms and performance analysis," IEEE Trans. Inf. Forensics Security, vol. 15, pp. 3454–3469, 2020.
11. 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. 102994, 2021.
12. T. M. Booij, I. Chiscop, E. Meeuwissen, N. Moustafa, and F. T. H. den Hartog, "ToN_IoT: The role of heterogeneity and the need for standardization of features and attack types in IoT network intrusion datasets," IEEE Internet Things J., vol. 9, no. 1, pp. 485–496, 2022.
13. 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.
14. D. C. Attota, V. Mothukuri, R. M. Parizi, and S. Pouriyeh, "An ensemble multi-view federated learning intrusion detection for IoT," IEEE Access, vol. 9, pp. 117734–117745, 2021.
15. 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. 108693, 2022.
16. O. Aouedi, K. Piamrat, G. Muller, and K. Singh, "Federated semisupervised learning for attack detection in industrial Internet of Things," IEEE Trans. Ind. Informat., vol. 19, no. 1, pp. 286–295, 2023.
17. M. Abdel-Basset, N. Moustafa, H. Hawash, I. Razzak, K. M. Sallam, and O. M. Elkomy, "Federated intrusion detection in blockchain-based smart transportation systems," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 3, pp. 2523–2537, 2022.
18. Z. Chen, N. Lv, P. Liu, Y. Fang, K. Chen, and W. Pan, "Intrusion detection for wireless edge networks based on federated learning," IEEE Access, vol. 8, pp. 217463–217472, 2020.
19. P. Ruzafa-Alcazar et al., "Intrusion detection based on privacy-preserving federated learning for the industrial IoT," IEEE Trans. Ind. Informat., vol. 19, no. 2, pp. 1145–1154, 2023.
20. M. Sarhan, S. Layeghy, N. Moustafa, and M. Portmann, "Cyber threat intelligence sharing scheme based on federated learning for network intrusion detection," J. Netw. Syst. Manage., vol. 31, no. 1, art. 3, 2023.
21. O. Belarbi, T. Spyridopoulos, E. Anthi, I. Mavromatis, P. Carnelli, and A. Khan, "Federated deep learning for intrusion detection in IoT networks," in Proc. IEEE GLOBECOM, 2023, pp. 237–242.
22. Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Differentially private asynchronous federated learning for mobile edge computing in urban informatics," IEEE Trans. Ind. Informat., vol. 16, no. 3, pp. 2134–2143, 2020.
23. D. Man, F. Zeng, W. Yang, M. Yu, J. Lv, and Y. Wang, "Intelligent intrusion detection based on federated learning for edge-assisted Internet of Things," Secur. Commun. Netw., vol. 2021, art. 9361348, 2021.
24. C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," Knowl.-Based Syst., vol. 216, art. 106775, 2021.