SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Federated Learning for Privacy-Preserving Cyber Threat Detection in Distributed IoT Networks

Kartikkumar Ashokbhai Pandya[1], Devarsh Mukeshbhai Sukhadiya[2]

[1]Student Researcher (Computer Science) Arizona State University, Arizona, United States.
[2]Software Engineer MABLS, Arizona, United States.

Abstract
This study examines federated learning as a privacy-conscious approach to cyber-threat detection in distributed Internet of Things (IoT) environments, where centralized processing may expose sensitive network information. The study aims to establish a structured federated-learning framework for cyber-threat detection and to examine detection effectiveness using standard classification measures while considering the influence of aggregation strategies and privacy preservation. The analysis draws on published benchmark evaluations using the ToN_IoT cybersecurity dataset. Reported results indicate that federated learning can maintain consistent cyber-threat detection performance across distributed clients, with the aggregated federated model achieving competitive performance compared with centralized learning. The comparison of FedAvg, FedAvgM, FedAdam, and FedAdagrad further demonstrates that the choice of aggregation mechanism can substantially influence classification performance. Published privacy-preserving federated-learning results also indicate that collaborative threat detection can be supported within a distributed data architecture while avoiding direct centralization of underlying network observations. These findings highlight the potential of federated learning for privacy-conscious cybersecurity in distributed IoT networks. However, broader evaluation involving heterogeneous clients, non-IID data distributions, resource-constrained devices, dynamically changing attack patterns, and stronger privacy mechanisms is required. Future validation across independent IoT cybersecurity datasets and real-world deployment environments would further strengthen evidence regarding the robustness, scalability, and generalizability of federated cyber-threat detection.

Keywords: Cybersecurity, Federated learning, Internet of things, Privacy preservation, Threat detection

1. Introduction

The rapid growth of the Internet of Things (IoT) has transformed conventional networks into highly interconnected environments involving sensors, smart devices, industrial systems, healthcare equipment, and cloud platforms. This connectivity supports automation, real-time monitoring, and intelligent decision-making, but it also creates a broader attack surface for cyber threats (Alazab *et al*., 2023). The heterogeneous nature of IoT devices, differences in computational capabilities, and continuous exchange of network information make conventional centralized security mechanisms increasingly challenging to maintain (Al Amro, 2025). Machine-learning-based intrusion detection has consequently become an important approach for identifying abnormal and malicious activities in IoT environments (Gutti *et al*., 2025). Recent studies have further demonstrated the applicability of federated learning to distributed intrusion detection while addressing the privacy concerns associated with centralized data collection (Raza *et al*., 2024).

Federated learning (FL) provides a decentralized learning paradigm in which multiple participating devices or organizations train models using their locally available data and share model updates rather than transferring raw information to a central repository. This characteristic makes FL particularly attractive for privacy-sensitive IoT cybersecurity applications (Ruzafa-Alcázar *et al*., 2021). By retaining security-related data at the local level, FL can reduce direct exposure of sensitive network information while enabling participating clients to contribute collectively to a global detection model. Recent research has explored this concept in IoT networks, wireless sensor networks, software-defined networks, and industrial environments (Bukhari *et al*., 2024).

Distributed IoT networks present several challenges that affect the reliability and scalability of cyber-threat detection. IoT devices generate data with different statistical characteristics, communication patterns, and attack profiles (Awan *et al*., 2023). Consequently, the resulting client datasets may be non-independent and non-identically distributed (non-IID), which can negatively influence collaborative model training. In addition, resource-constrained devices may have limited processing power, storage capacity, and communication

bandwidth. These constraints make it necessary to design security solutions that can operate efficiently without requiring continuous transmission of large volumes of raw data.

Privacy represents another major concern. Although FL avoids direct exchange of raw training data, model updates may still contain information that could potentially reveal characteristics of the underlying datasets. Therefore, federated intrusion-detection systems require additional mechanisms to strengthen confidentiality and protect distributed participants. Cryptographic techniques, secure aggregation, differential privacy, and blockchain-supported approaches have been investigated to address these concerns (Abou El Houda *et al*., 2023). Such approaches demonstrate that privacy protection must be considered alongside detection performance when designing collaborative cybersecurity systems.

The security problem becomes more complex when IoT networks contain multiple attack categories and heterogeneous sources. Distributed intrusion-detection systems must identify malicious behaviour accurately while minimizing false alarms and maintaining consistent performance across participating clients (Mahmud *et al*., 2024). Research involving wireless sensor networks and cyber-physical systems has therefore examined federated and hybrid deep-learning approaches for improving detection reliability. These developments highlight the need for flexible learning frameworks capable of accommodating distributed and heterogeneous cybersecurity data.

The application of FL to cybersecurity enables multiple network participants to collaboratively improve a detection model without requiring direct sharing of their raw traffic records (Nandanwar and Katarya, 2024). A typical FL process involves local model training followed by transmission of model parameters or updates to an aggregation server, where the contributions are combined to produce an updated global model. The resulting model is subsequently distributed to participating clients (Aramide, 2025). This architecture has been investigated for IoT intrusion detection and industrial IoT security, demonstrating the potential of decentralized learning for privacy-sensitive environments (Khraisat *et al*., 2025).

Several studies have extended this concept by integrating FL with additional security technologies. Blockchain and software-defined networking have been combined with FL to strengthen collaborative attack mitigation and trust management. Privacy-preserving mechanisms have also been investigated for consumer IoT and broader network environments (Rabieinejad *et al*., 2024). Other recent studies have examined FL-based cybersecurity frameworks for cloud systems, smart cities, intelligent transportation systems, and IoT-assisted infrastructures (Ragab et al., 2025).

The use of advanced learning architectures further demonstrates the growing interest in improving federated cyber-threat detection. Hybrid deep-learning models, including combinations of convolutional and recurrent architectures, have been investigated for distributed intrusion detection in wireless sensor networks and IoT environments (Jena *et al*., 2024). More recent work has considered scalable and privacy-preserving FL for IoT-enabled healthcare systems and distributed network-security applications (Alaskar *et al*., 2026). These studies collectively indicate that FL can provide a foundation for collaborative cybersecurity while reducing the requirement for centralized storage of sensitive network information.

Despite substantial progress, existing studies reveal several unresolved issues. Current FL-based cybersecurity solutions differ in their model architectures, aggregation mechanisms, privacy techniques, client configurations, and evaluation criteria (Al Amro, 2025). Some approaches primarily optimize intrusion-detection accuracy, whereas others emphasize privacy, blockchain-based trust, scalability, or real-time threat identification (Rahmati and Pagano, 2025). This variation makes it difficult to establish a consistent understanding of the trade-off between detection effectiveness and privacy protection in distributed IoT networks.

Furthermore, challenges associated with non-IID data, heterogeneous clients, communication overhead, and potentially malicious participants remain important considerations. Privacy-preserving FL frameworks have addressed some of these issues, but further investigation is required to develop approaches that simultaneously support effective cyber-threat detection and protection of distributed data (Mankotia *et al*., 2026). Recent work on privacy-preserving intrusion detection, smart-city cybersecurity, and cloud-based threat detection reinforces the need for adaptable FL architectures capable of operating across diverse IoT environments (Nagib *et al*., 2025).

Accordingly, the research gap lies in establishing a unified perspective that treats distributed cyber-threat detection and privacy preservation as interconnected requirements rather than independent objectives. The present study addresses this gap by focusing on federated learning as a collaborative mechanism for cyber-threat detection in distributed IoT networks while considering the privacy requirements associated with locally retained security data. The approach is consistent with previous evidence demonstrating the suitability of privacy-preserving FL for industrial IoT and related distributed environments (Moulahi *et al*., 2023).

This study has two primary objectives. First, it aims to establish a federated-learning framework for cyber-threat detection in distributed IoT networks while maintaining the privacy of locally generated network data.

Second, it aims to evaluate the effectiveness of the federated approach using standard detection measures, including accuracy, precision, recall, and F1-score, while examining the role of federated aggregation and privacy preservation in collaborative cybersecurity.

2. Materials and Methods

2.1 Research Framework

This study adopts a secondary-data research design to investigate privacy-preserving federated learning for cyber-threat detection in distributed IoT networks. The analysis considers the ToN_IoT dataset, developed by UNSW Canberra Cyber for evaluating artificial-intelligence and machine learning techniques in IoT, Industrial IoT, and Industry 4.0 cybersecurity environments. The dataset contains heterogeneous telemetry and network-security information representing normal and malicious activities, together with corresponding security-event ground truth. As an existing publicly available dataset, ToN_IoT provides a suitable data foundation for investigating federated cyber-threat detection without requiring the collection of new information directly from IoT devices or network participants.

2.2 Data Preparation

The analysis is based on secondary ToN_IoT data and considers the preprocessing requirements associated with preparing the dataset for machine-learning-based cyber-threat detection. Relevant records and features are considered in accordance with the published evaluations on which the reported results are based. Data preparation may include the treatment of unsuitable observations, numerical representation of categorical variables, and feature scaling where required by the underlying learning framework. The target variable is defined according to the cyber-threat detection task. These preprocessing considerations provide the basis for organizing the available observations for distributed learning without implying the collection of new IoT network data.

2.3 Distributed IoT Client Configuration

The distributed-learning setting is represented through virtual IoT clients, consistent with the federated-learning configuration reported in the underlying ToN_IoT evaluation. Each virtual client represents an independent participant within a distributed IoT environment and is associated with a separate portion of the available observations. Raw observations are retained at the client level rather than exchanged directly between participants. A common cyber-threat detection model can then be considered across the participating clients, allowing local information to contribute to collaborative model development. This configuration provides a basis for examining federated learning in environments where cybersecurity information originates from distributed IoT participants.

2.4 Local Training and Federated Aggregation

The federated-learning process considered in this study follows the configuration reported in the published ToN_IoT evaluation. In this configuration, a global model is distributed to participating virtual clients, where local model training is performed using the respective client data. Rather than transmitting raw observations, model parameters or updates are communicated to an aggregation server. The reported evaluation uses Federated Averaging (FedAvg) to combine the client updates and obtain a global model. The aggregated model can subsequently be evaluated against the reported client-level and centralized results. This procedure illustrates how collaborative learning can be performed without directly pooling the underlying network records at a central location.

2.5 Privacy-Preserving Cyber-Threat Detection

Privacy protection in the federated-learning setting is primarily considered through local retention of the underlying network observations. Instead of directly transferring raw cybersecurity records to a central server, the federated architecture communicates model parameters or updates required for collaborative aggregation. The detection task focuses on distinguishing normal and malicious network behaviour using the available ToN_IoT features and security labels. Detection effectiveness is assessed using accuracy, precision, recall, and F1-score. This privacy-aware configuration is consistent with previous research applying federated learning to privacy-sensitive intrusion detection in IoT and industrial IoT environments.

2.6 Experimental Evaluation

The evaluation considered in this study is based on reported experimental results from published ToN_IoT evaluations. The analysis examines federated client-level performance, aggregated federated performance, and centralized-learning performance to assess the effectiveness of distributed cyber-threat detection. Published

results for alternative federated aggregation strategies are also considered to examine their influence on classification performance. Accuracy, precision, recall, and F1-score are used as the principal detection measures, while false-positive rate is considered where reported. The evaluation is therefore based on existing secondary data and published experimental evidence rather than newly collected IoT observations. The dataset characteristics, federated-learning configuration, aggregation procedures, and evaluation criteria are considered to provide methodological context for interpreting the reported findings.

3. Results

3.1 Federated Detection Capability Across Distributed Clients

Published results demonstrate the effectiveness of federated learning for cyber-threat detection across distributed IoT clients. The four federated clients reported closely comparable performance, with accuracy values ranging from 97.92% to 98.13%. The aggregated federated model reported 97.86% accuracy and an F1-score of 97.85%, indicating consistent classification performance after combining knowledge from distributed participants, as shown in Table 1. Precision and recall for the aggregated model were 97.89% and 97.86%, respectively. Although the centralized model reported a higher accuracy of 99.40%, the relatively small difference indicates that federated learning can achieve competitive detection performance while maintaining a distributed training structure. These published findings support the potential of federated learning for IoT cyber-threat detection.

Table 1. Federated Learning Performance for IoT Cyber-Threat Detection on ToN_IoT

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
FL Client 1	98.13	98.16	98.13	98.12
FL Client 2	98.06	98.10	98.06	98.06
FL Client 3	97.92	97.94	97.92	97.91
FL Client 4	98.00	98.03	98.00	98.00
Aggregated FL Model	97.86	97.89	97.86	97.85
Centralized Model	99.40	99.40	99.40	99.40

Source: *Lazzarini, Tianfield, and Charissis (2023),* Federated Learning for IoT Intrusion Detection*, AI, 4, 509–530. Reported ToN_IoT experimental results.*

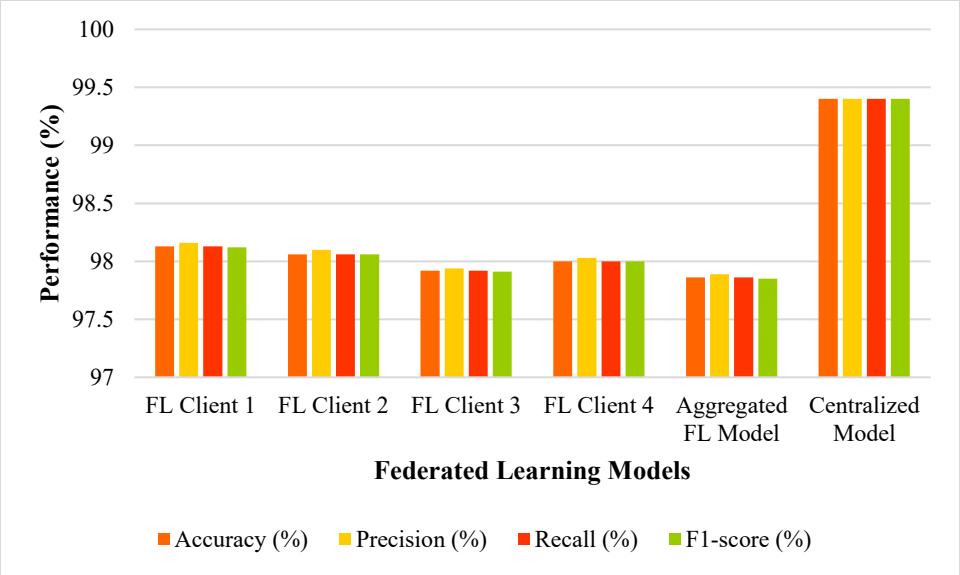

Figure 1. Comparative Cyber-Threat Detection Performance of Federated and Centralized Learning Models

The reported detection performance of federated learning clients compared with a centralized learning model is presented in Figure 1. The four federated clients demonstrate highly consistent reported results, with accuracy, precision, recall, and F1-score remaining close to 98%. The aggregated federated model also maintains strong performance, although its measures are slightly lower than those of the individual clients. In contrast, the centralized model reports approximately 99.4% across all four evaluation measures. These

published results indicate that federated learning can achieve competitive cyber-threat detection performance while supporting a distributed learning architecture. The relatively small performance difference highlights its potential for privacy-conscious IoT cybersecurity applications.

3.2 Influence of Federated Aggregation Strategy on Detection Performance

The reported performance of different federated aggregation strategies for IoT cyber-threat detection demonstrates the influence of the aggregation mechanism on classification performance. FedAvg reported the strongest overall performance, with 97.86% accuracy, 97.89% precision, 97.86% recall, and 97.85% F1-score as shown in Table 2. FedAvgM produced comparable performance, reporting 97.57% accuracy, 97.66% precision, 97.57% recall, and 97.58% F1-score. In contrast, FedAdam and FedAdagrad reported lower accuracy and F1-score values. These published results indicate that the aggregation strategy is an important factor in federated intrusion-detection systems. Consequently, selecting an appropriate aggregation mechanism can contribute to maintaining reliable threat-detection performance across distributed IoT environments.

Table 2. Performance of Federated Aggregation Algorithms on ToN_IoT

Federated Learning Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
FedAvg	97.86	97.89	97.86	97.85
FedAvgM	97.72	95.12	98.53	96.79
FedAdam	87.67	75.80	95.05	84.34
FedAdagrad	81.76	66.35	96.97	78.79

Source: *Lazzarini, Tianfield, and Charissis,* Federated Learning for IoT Intrusion Detection*, AI, 2023. The study evaluates FedAvg, FedAvgM, FedAdam, and FedAdagrad in the same federated-learning setting.*

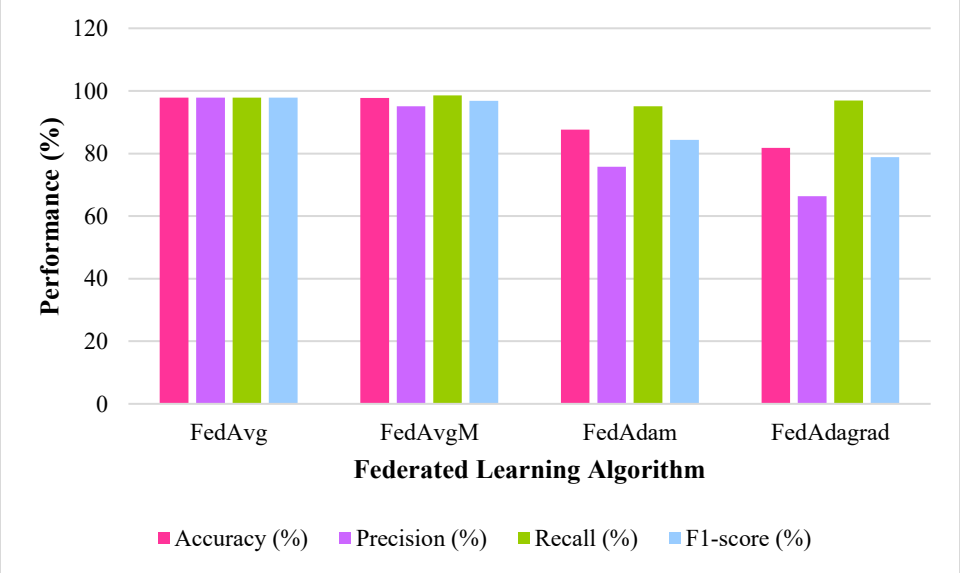

Figure 2. Performance Comparison of Federated Learning Aggregation Algorithms

The reported performance of four federated learning aggregation algorithms for IoT cyber-threat detection. FedAvg demonstrates strong and balanced performance across all evaluation measures, with accuracy, precision, recall, and F1-score remaining close to 98%, as shown in Figure 2. FedAvgM produces comparable performance across the four measures, with accuracy, precision, recall, and F1-score of 97.57%, 97.66%, 97.57%, and 97.58%, respectively. In comparison, FedAdam shows a noticeable reduction in accuracy, recall, and F1-score, although its precision remains comparatively higher than its accuracy and recall. FedAdagrad records the lowest accuracy, precision, recall, and F1-score among the evaluated approaches. The comparison indicates that the choice of aggregation algorithm can substantially influence federated cyber-threat detection performance in distributed IoT environments.

3.3 Benchmarking Privacy-Preserving Federated Learning

The published benchmark results are from a privacy-preserving federated-learning framework evaluated on the ToN_IoT dataset. The framework reported an accuracy of 89.60%, accompanied by precision of 88.70% and

recall of 85.30%. Its F1-score was 86.10%, indicating a balance between precision and recall, as shown in Table 3. The reported false-positive rate of 3.10% is particularly relevant to practical IoT security environments, where excessive false alarms may reduce the usefulness of automated detection systems. Together, these reported measures demonstrate the potential of privacy-aware federated learning to support cyber-threat detection while maintaining a distributed learning paradigm for data-sensitive IoT infrastructures.

Table 3. Privacy-Preserving Federated Learning Performance on ToN_IoT

Performance Metric	ToN_IoT Result (%)
Accuracy	89.60
Precision	88.70
Recall	85.30
F1-score	86.10
False Positive Rate	3.10

Source: *Alqazzaz,* SecuFL-IoT: an adaptive privacy-preserving federated learning framework for anomaly detection in smart industrial networks*, Scientific Reports, 2026. The paper explicitly reports these results from its cross-dataset validation on ToN_IoT.*

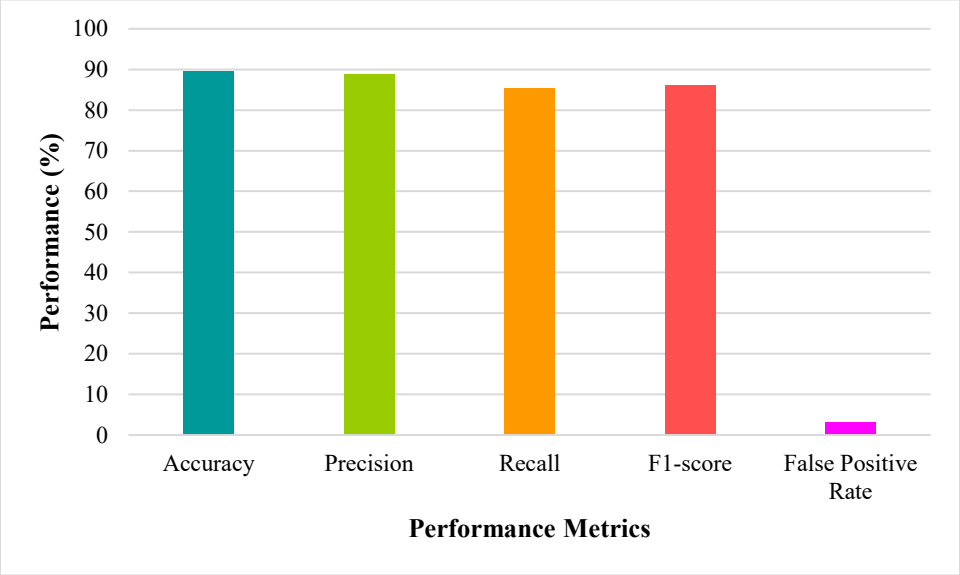

Figure 3. Privacy-Preserving Federated Learning Performance

The reported performance of a privacy-preserving federated-learning framework evaluated on the ToN_IoT dataset. The framework reported an accuracy of 89.60% and precision of 88.70%, indicating effective classification performance, as shown in Figure 3. The reported recall of 85.30% indicates that most malicious instances were identified, although some threats may remain undetected. The F1-score of 86.10% reflects a balance between precision and recall. Notably, the reported false-positive rate was 3.10%, indicating a relatively low rate of legitimate network activity being incorrectly classified as malicious. Overall, these reported results demonstrate the potential of privacy-preserving federated learning for distributed IoT cybersecurity applications.

4. Discussion

The published benchmark findings indicate that federated learning can provide strong cyber-threat detection performance in distributed Internet of Things (IoT) environments while reducing the need for direct centralization of network observations. The results summarized in Table 1 show that the four federated clients achieved highly consistent performance, with accuracy values ranging from 97.92% to 98.13%. The aggregated federated model reported 97.86% accuracy and a 97.85% F1-score, while precision and recall were 97.89% and 97.86%, respectively (Lazzarini *et al*., 2023). Although the centralized model reported 99.40% across the four evaluation measures, the relatively small performance difference suggests that federated learning can provide competitive detection performance in scenarios where data localization and privacy requirements are important. These findings are consistent with the broader use of federated learning as a distributed approach

for cybersecurity applications in which sensitive data may remain at participating organizations or devices rather than being directly pooled in a central repository (Vyas *et al*., 2024).

The client-level results further indicate that federated learning can maintain relatively stable detection performance when available observations are distributed among multiple participating clients. The differences between the four clients are small, suggesting that the evaluated federated configuration was able to learn relatively consistent representations of the cyber-threat patterns present in the underlying dataset (Lazzarini *et al*., 2023). The aggregated model reported slightly lower performance than the individual client models, which may reflect the challenges associated with combining information originating from different distributed environments. Nevertheless, the aggregated model maintained accuracy, precision, recall, and F1-score close to 98%. This consistency is important for distributed IoT security because substantial variation in detection capability across participating nodes could reduce the reliability of a collaborative intrusion-detection system (Timofte *et al*., 2025). Previous research similarly emphasizes the importance of collaborative learning for intrusion detection in IoT environments while recognizing the challenges associated with heterogeneous and distributed data (Vyas *et al*., 2024).

Privacy preservation is a central consideration in distributed IoT cybersecurity because network traffic and security-related telemetry may contain sensitive operational information. The benchmark summarized in Table 3 reported 89.60% accuracy, 88.70% precision, 85.30% recall, and 86.10% F1-score, together with a reported false-positive rate of 3.10% (Alqazzaz, 2026). These results were reported for cross-dataset validation on ToN_IoT and therefore provide supporting benchmark evidence rather than newly generated results from the present study. The reported values indicate that the evaluated privacy-aware federated-learning framework retained useful detection capability while reducing the need to directly centralize the underlying observations. The relatively low false-positive rate is particularly relevant to practical IoT security environments because excessive false alarms may increase the workload of security administrators and reduce confidence in automated detection systems. Accordingly, privacy preservation should be considered together with detection effectiveness rather than treated as an independent design objective.

The comparison between centralized and federated learning reveals a performance advantage for centralized training in the reported benchmark, with the centralized model achieving 99.40% across accuracy, precision, recall, and F1-score (Lazzarini *et al*., 2023). However, higher centralized performance does not necessarily make centralized learning preferable for every IoT cybersecurity scenario. Centralized approaches generally require security-related observations to be transferred to a common processing location, whereas federated learning allows participating entities to contribute to collaborative model development without directly pooling their underlying records. Federated learning therefore provides a different privacy and data-governance trade-off rather than simply representing a higher-performing alternative to centralized learning. The relatively limited performance difference observed in the reported ToN_IoT evaluation can consequently be interpreted as a potential trade-off between maximum predictive performance and data-localization requirements. For sensitive IoT environments, accepting a modest reduction in predictive performance may be justified when retaining data locally is an important operational or privacy requirement. Similar motivations have been identified in research examining federated learning for privacy-preserving intrusion detection and cybersecurity (Timofte *et al*., 2025).

The comparison of federated aggregation strategies also demonstrates that the choice of aggregation mechanism can substantially influence detection performance. In the reported multiclass ToN_IoT evaluation, FedAvg achieved 97.86% accuracy, 97.89% precision, 97.86% recall, and 97.85% F1-score, whereas FedAvgM reported 97.57%, 97.66%, 97.57%, and 97.58%, respectively. FedAdam and FedAdagrad produced considerably lower performance across several measures (Lazzarini et al., 2023). These differences indicate that federated learning performance depends not only on the decision to distribute training but also on how local model updates are aggregated. Consequently, selecting an appropriate aggregation strategy is an important design consideration for federated intrusion-detection systems. The finding is consistent with the broader literature, which identifies aggregation, client heterogeneity, data imbalance, and communication efficiency as important factors affecting federated-learning performance in cybersecurity applications (Tabassum *et al*., 2022). In particular, the FEDGAN-IDS framework demonstrates that federated intrusion detection can also be combined with data-generation techniques to address challenges associated with insufficient or imbalanced training data.

The findings have practical implications for IoT security architectures involving numerous interconnected devices, organizations, and network participants. The high client-level performance reported in the underlying evaluation suggests that collaborative learning can support cyber-threat detection across distributed environments without requiring every participant to transfer its complete locally maintained observations to a central repository. This characteristic is particularly relevant to IoT ecosystems in which data may be generated across geographically distributed devices and administrative domains. Privacy-preserving

federated-learning research has emphasized the potential of keeping training data local while sharing model information for collaborative detection. More recent cybersecurity research has also explored federated-learning architectures that combine privacy preservation with robustness, indicating that practical deployment requires consideration of threats against both the protected data and the collaborative learning process (Soomro *et al*., 2025). Therefore, the practical value of federated learning extends beyond classification accuracy to include data governance, privacy requirements, and the operational constraints of distributed IoT environments.

Federated learning also offers a potentially scalable structure because additional participants can contribute local model updates without requiring their complete datasets to be transferred to a central repository. However, scalability should not be interpreted solely in terms of the number of participating clients. Increasing participation may introduce differences in data distributions, computational capabilities, communication conditions, device resources, and attack frequencies. Consequently, the strong results reported for the available client configuration cannot automatically be generalized to large-scale or highly heterogeneous IoT deployments. The literature identifies non-IID data, communication efficiency, resource constraints, and client heterogeneity as continuing challenges for federated cybersecurity systems. Broader validation involving more heterogeneous clients, non-IID data distributions, varying device resources, and changing attack patterns would therefore provide stronger evidence concerning the robustness and generalizability of federated cyber-threat detection.

The findings are also consistent with the broader direction of research combining federated learning, privacy preservation, and cybersecurity. Tabassum et al. (2022) proposed FEDGAN-IDS, which combines federated learning and generative adversarial networks for privacy-preserving intrusion detection in smart IoT environments and highlights the importance of addressing data imbalance within distributed learning. The reviewed privacy-preserving federated learning for intrusion detection in IoT environments and emphasized the role of mechanisms such as federated learning, differential privacy, homomorphic encryption, and secure multiparty computation in strengthening privacy protection. Timofte *et al*. (2025) further investigated federated learning as a privacy-preserving cybersecurity approach and highlighted challenges associated with distributed cybersecurity learning. More recent work has extended this direction toward robust privacy-preserving federated intrusion detection for IoT and industrial IoT environments. In addition, Yazdinejad *et al*. (2025) incorporated explainability into privacy-preserving federated threat detection for cyber-physical-social systems, demonstrating the increasing importance of interpretability alongside privacy and detection performance. Collectively, these studies reinforce the relevance of integrating collaborative learning, privacy protection, robustness, and explainability within distributed cybersecurity architectures.

Several limitations should be acknowledged when interpreting the reported findings. First, the present analysis relies on secondary benchmark data and published experimental evidence rather than newly collected IoT network traffic. Consequently, the reported results may not fully represent contemporary operational IoT environments with continuously changing devices, protocols, traffic patterns, and attack behaviours. Second, the numerical results originate from existing experimental studies and may therefore involve methodological configurations that differ in preprocessing procedures, client partitioning, model architecture, hyperparameter selection, and evaluation protocols. Direct numerical comparison across studies should consequently be interpreted with caution. Third, privacy preservation in the present analysis is considered primarily from the perspective of keeping underlying observations local and communicating model information rather than raw records. Basic federated learning should not be interpreted as providing complete privacy protection because model updates may themselves create privacy risks. Stronger privacy guarantees may require additional mechanisms, including secure aggregation, differential privacy, encryption, or protection against inference and model-poisoning attacks. Finally, the limited client configuration represented in the underlying benchmark does not fully capture the complexity of large-scale, heterogeneous IoT deployments.

Future research should therefore evaluate federated cyber-threat detection under larger and more heterogeneous client populations, non-IID data distributions, and dynamically changing attack patterns. Additional privacy mechanisms should be investigated to strengthen protection against inference attacks and potential information leakage from model updates. Research should also examine communication overhead, computational requirements, energy consumption, and convergence behaviour to establish the practical feasibility of federated-learning deployment on resource-constrained IoT devices. Robust aggregation mechanisms and defenses against malicious or compromised participants should receive further attention because distributed cybersecurity systems must protect not only the underlying data but also the collaborative learning process. Explainable artificial intelligence could further improve the transparency and interpretability of federated threat-detection decisions, particularly in security environments where analysts require understandable evidence for automated alerts. Finally, evaluation across multiple independent IoT cybersecurity datasets and real-world heterogeneous environments would strengthen the generalizability of

the findings and provide a more comprehensive assessment of privacy-preserving federated learning for distributed IoT cyber-threat detection.

The published benchmark evidence suggests that federated learning can achieve competitive cyber-threat detection performance while providing a data-localization-oriented alternative to centralized learning. The results also demonstrate that aggregation strategy, client heterogeneity, privacy mechanisms, and deployment constraints can substantially influence the practical effectiveness of federated cybersecurity systems. Therefore, federated learning should not be viewed simply as a replacement for centralized intrusion detection, but rather as a distributed learning paradigm whose advantages must be evaluated jointly in terms of detection effectiveness, privacy protection, robustness, scalability, and operational feasibility.

5. Conclusion

This study provides a structured assessment of federated learning for privacy-conscious cyber-threat detection in distributed Internet of Things (IoT) networks. The reported benchmark evidence indicates that collaborative learning can achieve competitive detection performance while allowing participating environments to retain their underlying network observations locally. The analysis also highlights the influence of federated aggregation mechanisms on detection effectiveness, demonstrating that the selection of an appropriate aggregation strategy is an important consideration when designing distributed cybersecurity systems. The findings emphasize that federated learning can provide an alternative to conventional centralized approaches when direct sharing of security-related data is undesirable or impractical. By supporting collaborative model development without requiring the transfer of raw network observations, the approach can contribute to privacy-aware security monitoring across distributed IoT infrastructures. However, local data retention alone should not be interpreted as complete privacy protection, as additional safeguards may be required to protect model updates from potential information leakage and inference attacks. Future research should examine federated threat detection across larger and more heterogeneous client populations, non-IID data distributions, resource-constrained devices, and dynamically evolving attack behaviours. Communication overhead, computational and energy requirements, robust aggregation, stronger privacy mechanisms, and resilience against malicious participants should also receive greater attention. Finally, validation using multiple independent IoT cybersecurity datasets and real-world deployment environments would strengthen evidence regarding the robustness, scalability, and generalizability of privacy-preserving federated cyber-threat detection.

References

1. Abou El Houda, Z., Hafid, A. S., & Khoukhi, L. (2023). MiTFed: A privacy-preserving collaborative network attack mitigation framework based on federated learning using SDN and blockchain. *IEEE Transactions on Network Science and Engineering*, *10*(4), 1985-2001.
2. Al Amro, S. (2025). Securing Internet of Things devices with federated learning: A privacy-preserving approach for distributed intrusion detection. *Computers, Materials & Continua*, *83*(3), 4623-4658.
3. Alaskar, N. M., Hussain, M., Almheiri, S. J., Khan, A., & Adnan, K. M. (2026). Big Data-Driven Federated Learning Model for Scalable and Privacy-Preserving Cyber Threat Detection in IoT-Enabled Healthcare Systems. *Computers, Materials & Continua*, *87*(1), 29.
4. Alazab, A., Khraisat, A., Singh, S., & Jan, T. (2023). Enhancing privacy-preserving intrusion detection through federated learning. *Electronics*, *12*(16), 3382.
5. Alqazzaz, A. (2026). SecuFL-IoT: An adaptive privacy-preserving federated learning framework for anomaly detection in smart industrial networks. *Scientific Reports*, *16*(1), 4107.
6. Aramide, O. O. (2025). Federated learning for distributed network security and threat intelligence: A privacy-preserving paradigm for scalable cyber defense. *Journal of Data Analysis and Critical Management*, *1*(02).
7. Awan, K. A., Din, I. U., Almogren, A., & Rodrigues, J. J. (2023). Privacy-preserving big data security for IoT with federated learning and cryptography. *IEEE access*, *11*, 120918-120934.
8. Bukhari, S. M. S., Zafar, M. H., Abou Houran, M., Moosavi, S. K. R., Mansoor, M., Muaaz, M., & Sanfilippo, F. (2024). Secure and privacy-preserving intrusion detection in wireless sensor networks: Federated learning with SCNN-Bi-LSTM for enhanced reliability. *Ad Hoc Networks*, *155*, 103407.
9. Gutti, C., Thumula, K., & Balbudhe, P. (2025). Federated learning for distributed IoT security: A privacy-preserving approach to intrusion detection. *IEEE Access*, *13*, 135863-135875.
10. Jena, S. R., Rahman, M. Z. U., Sinha, D. K., Kumar, P. R., Vimal, V., Singh, K. U., ... & Balajee, J. (2024). An innovative secure and privacy-preserving federated learning-based hybrid deep learning model for intrusion detection in internet-enabled wireless sensor networks. *IEEE Transactions on Consumer Electronics*, *71*(1), 273-280.

11. Khraisat, A., Alazab, A., Alazab, M., Obeidat, A., Singh, S., & Jan, T. (2025). Federated learning for intrusion detection in IoT environments: a privacy-preserving strategy. *Discover Internet of Things*, *5*(1), 72.
12. Lazzarini, R., Tianfield, H., & Charissis, V. (2023). Federated learning for IoT intrusion detection. *Ai*, *4*(3), 509-530.
13. Mahmud, S. A., Islam, N., Islam, Z., Rahman, Z., & Mehedi, S. T. (2024). Privacy-preserving federated learning-based intrusion detection technique for cyber-physical systems. *Mathematics*, *12*(20), 3194.
14. Mankotia, S., Conte de Leon, D., & Rimal, B. P. (2026). FedPrIDS: Privacy-preserving federated learning for collaborative network intrusion detection in IoT. *Journal of cybersecurity and privacy*, *6*(1), 10.
15. Moulahi, T., Jabbar, R., Alabdulatif, A., Abbas, S., El Khediri, S., Zidi, S., & Rizwan, M. (2023). Privacy-preserving federated learning cyber-threat detection for intelligent transport systems with blockchain-based security. *Expert Systems*, *40*(5), e13103.
16. Nagib, M. N., Pervez, R., Nova, A. A., Ahmed, M., Maua, J., & Sayeema, A. (2025, July). FedSAT-Detect: A Federated Learning Framework for Privacy-Preserving Cybersecurity Threat Detection in Cloud Systems. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE.
17. Nandanwar, H., & Katarya, R. (2024, December). A secure and privacy-preserving ids for iot networks using hybrid blockchain and federated learning. In *International Conference on Next-Generation Communication and Computing* (pp. 207-219). Singapore: Springer Nature Singapore.
18. Rabieinejad, E., Yazdinejad, A., Dehghantanha, A., & Srivastava, G. (2024). Two-level privacy-preserving framework: Federated learning for attack detection in the consumer Internet of Things. *IEEE Transactions on Consumer Electronics*, *70*(1), 4258-4265.
19. Ragab, M., Ashary, E. B., Alghamdi, B. M., Aboalela, R., Alsaadi, N., Maghrabi, L. A., & Allehaibi, K. H. (2025). Advanced artificial intelligence with a federated learning framework for privacy-preserving cyberthreat detection in IoT-assisted sustainable smart cities. *Scientific reports*, *15*(1), 4470.
20. Rahmati, M., & Pagano, A. (2025, July). Federated learning-driven cybersecurity framework for IoT networks with privacy preserving and real-time threat detection capabilities. In *Informatics* (Vol. 12, No. 3, p. 62). MDPI.
21. Raza, M., Saeed, M. J., Riaz, M. B., & Sattar, M. A. (2024). Federated learning for privacy-preserving intrusion detection in software-defined networks. *Ieee Access*, *12*, 69551-69567.
22. Ruzafa-Alcázar, P., Fernández-Saura, P., Mármol-Campos, E., González-Vidal, A., Hernández-Ramos, J. L., Bernal-Bernabe, J., & Skarmeta, A. F. (2021). Intrusion detection based on privacy-preserving federated learning for the industrial IoT. *IEEE Transactions on Industrial Informatics*, *19*(2), 1145-1154.
23. Soomro, I. A., Khan, H. U. R., Hussain, S. J., Iqbal, A., Khalid, W., & Yu, H. (2025). SecureDyn-FL: A robust privacy-preserving federated learning framework for intrusion detection in IoT networks. *IEEE Transactions on Network and Service Management*, *23*, 1742-1765.
24. Tabassum, A., Erbad, A., Lebda, W., Mohamed, A., & Guizani, M. (2022). Fedgan-ids: Privacy-preserving ids using gan and federated learning. *Computer Communications*, *192*, 299-310.
25. Timofte, E. M., Dimian, M., Graur, A., Potorac, A. D., Balan, D., Croitoru, I., ... & Puşcaşu, M. (2025). Federated learning for cybersecurity: A privacy-preserving approach. *Applied Sciences*, *15*(12), 6878.
26. Vyas, A., Lin, P. C., Hwang, R. H., & Tripathi, M. (2024). Privacy-preserving federated learning for intrusion detection in IoT environments: A survey. *IEEE access*, *12*, 127018-127050.
27. Yazdinejad, A., Mohammadabadi, Z. D., Dehghantanha, A., & Srivastava, G. (2025). An explainable and privacy-preserving federated learning model for threat detection in cyber-physical-social systems. *IEEE Transactions on Computational Social Systems*.