WFL: Edge-Enabled Weighted Federated Learning for Securing Heterogenous Networks

Muhammad Ajmal Azad, Ali Kashif Bashir, R. Muhammad Atif Azad and Syed Attique Shah

***Abstract*—The operational aspects of the Internet of Things (IoT) are dependent on the security measures deployed to ensure user privacy, protect user data and prevent smart devices from being exploited for malicious activities. Traditional Intrusion Detection Systems often require collaboration from many individual devices in the centralised system for data processing and decision-making. However, centralised systems have some limitations in terms of privacy and scalability. This paper proposes a federated learning-based (FL) distributed framework for detecting and mitigating intrusion while ensuring privacy in IoT networks. The framework integrates two key security components: an intrusion detection module that employs Neural Networks (NN) at the edge device, and centralised aggregation systems that aggregate and coordinate the aggregated model to edge devices. The centralised system computes the global model using a weighted averaging mechanism to accurately represent the relative importance of each device's local model. of each device's contribution. This ensures that the global model is the complete representation of the overall data at the collaborating edge nodes. The framework ensures privacy as data remains local to edge devices, and the machine learning models are exchanged to the aggregation server. By supporting heterogeneous data from various sources, the framework demonstrates adaptability to diverse attack patterns and device behaviours. The evaluation is conducted on heterogeneous datasets, including CICIDS2017, UNSW-NB15, and KDD Cup 99 under heterogeneous scenarios, which represent a wide range of intrusion scenarios, such as DDoS, Botnet activities and malicious behaviours. With an increased number of iterations and collaborators, the framework demonstrates improved performance, achieving an average intrusion detection accuracy of 99% across the three datasets. These results highlight the importance of both the number of collaborators and iterations in improving the overall model performance while preserving privacy and minimising communication overhead.**

***Index Terms*—Federated Learning, Collaborative Intrusion Detection, Privacy-preservation, Internet of Things**

I. Introduction

The Internet of Things (IoT) has emerged as a transformative technology as it brings drastic changes in various sectors such as healthcare, transportation, agriculture, and industrial automation. By 2025, the IoT market is projected to exceed $1.5 trillion and billions of IoT devices worldwide. However, the rapid growth of IoT networks has also exposed these networks to cyber threats, which could bring huge financial losses. Incidents such as the Mirai botnet [1] and the Ukrainian power grid cyberattack are just examples to show the impact of cyber attacks using IoT devices or attacks on IoT networks. As IoT devices continue to grow there is a need for effective and collaborative security systems to protect the networks from unwanted events.

Intrusion Detection Systems (IDS) have been widely used to secure heterogeneous networks from malicious events. IDS systems can employ signatures or ML models to identify malicious events. Signature-based detection compares traffic against known attack signatures and can achieve high accuracy but fails to identify unknown or zero-day attacks. Anomaly or ML-based detection systems identify deviations from the established baseline model. These systems can detect novel threats but often produce false positives due to their sensitivity to normal variations. Hybrid methods combine these approaches for improved coverage but add complexity and computational overhead. ML-based IDS systems can achieve high accuracy in heterogeneous systems while aggregating and processing the raw data from different devices. However, these systems process, aggregate and apply Machine learning in the centralised setup. Although the centralised systems achieve higher accuracy, they create a single point of failure, a single point of unauthorized access, and a potential threat to users' privacy as trusted systems can misuse sensitive shared data. Furthermore, the trusted system could use data without user consent. As a result, privacy-preserving mechanisms are needed that can achieve the trade-off between detection performance and privacy.

Federated Learning (FL) can be used to achieve higher accuracy without affecting the privacy of users and data contributors. FL enables collaborators to process and apply Machine Learning (ML) models to the local data at the edge or device level. The learned machine learning model is then shared with the centralised system instead of the raw data. In this scenario, privacy is achieved as data remains on the local device and only model updates are shared with centralised systems. FL also reduces communication overhead by transmitting only model updates, which are only a few hundred bytes rather than raw data. However, the effectiveness of FL depends on the quality of data used for training at the local device. In real-world scenarios, devices may have varying quality data, as some devices might have high-quality data, while others might create their model on noisy or irrelevant data. This heterogeneity of data might impact the overall performance of the global model.

In this paper, we propose a novel privacy-preserving edge-enabled weighted FL framework (WFL) to detect intrusions in IoT networks. The proposed approach uses the privacy-preserving properties of FL and introduces a weighted aggregation mechanism to compute the aggregated model. The weights assigned to each collaborator would help to consider

Muhammad Ajmal Azad, R. Muhammad Atif Azad and Syed Attqiue Shah are with the School of Computer Science, Birmingham City University, Ali Kashif Bashir is with Manchester Metropolitan University UK.

Manuscript received April 19, 2005; revised August 26, 2015.

the quality of the model contributed by different collaborators. Specifically, the system employs Neural Networks (NN) to train local models at the edge device and uses weighted averaging for global model aggregation. The weights assigned to each device reflect the trustworthiness and relevance of its local model. This ensures that the global model would be influenced more by high-quality and trustworthy updates, and malicious collaborators might not influence the performance of the system.

The major contributions of this research are as follows.

- We propose a new edge-enabled FL-based IDS framework that enables models to be trained at the edge device, and collaboration occurs across the collaborators. This method guarantees that data remains either on the local device or within the collaborators' operating jurisdiction. It achieves high detection accuracy and an acceptable false positive rate while fully protecting the data privacy of collaborators.
- We introduce a weighted aggregation mechanism that assigns different weights to collaborators. These weights have been computed according to the quality of relevance of local data used to train the model. This process would ensure that the global model is influenced more by high-quality updates and improve the overall accuracy and reliability of the intrusion detection system under heterogeneous networks.
- We conduct an extensive evaluation of the proposed system using three benchmark datasets: CICIDS2017, UNSW-NB15, and KDD Cup 99. Our experiment results demonstrate the efficacy of the proposed approach in comparison to traditional centralised and self-learning models.

The paper is organised as follows. Section II provides a detailed review of existing research. Section III explains the proposed methodology, including key technical approaches, models, and techniques used. The experimental setup and evaluation metrics are mentioned in section IV-A. Section V discusses the evaluation results across different performance metrics. Section VI concludes the paper with some future directions.

II. Related Works

In terms of deployment, IDS can be broadly classified into two categories: Host-Based IDS (HIDS) and Network-Based IDS (NIDS). HIDS are deployed on individual devices and monitor system-related activities, such as application files and operating system operations. In contrast, NIDS analyses network traffic by capturing and inspecting packet flows, making it more suitable for defending against external attacks. In terms of detection approach, IDS systems can be categorised into signature-based detection and anomaly-based detection systems [2], [3]. Within anomaly-based detection systems, machine learning and Artificial Intelligence have been widely used to identify intrusions within the network. Midi et al. [4] proposed an IDS for IoT networks that is not limited to specific protocols or applications. The system dynamically adapts its detection strategies based on network characteristics. Hodo et al. [5] employed a Multi-Layer Perceptron for an offline IoT intrusion detection system (IDS). Their approach analyses Internet packet traces to identify DoS and DDoS attacks within IoT networks. However, this method analysed traffic in an offline setting, which makes it unsuitable for real-time detection and collaboration. Moustafa et al. [6] adopted an ensemble-based network intrusion detection technique that uses statistical flow features to classify malicious activities. Al-Yaseen et al. [7] proposed a modified K-means algorithm to optimise the training dataset by reducing its size and balancing the data. This approach was used to train Support Vector Machines and Extreme Learning Machines. These systems normally require access to the whole data, which might be a threat to user privacy if not properly protected.

FL systems address the challenge of privacy preservation by enabling machine learning models to be trained locally on devices while collaborating through model sharing to a centralised or distributed aggregation system [8], [9], [10]. FL allows multiple distributed devices to collaboratively train a shared model without exposing raw data. The system can ensure privacy by using cryptographic systems, controlled noise addition, or differential privacy [11], [12]. However, these privacy-preserving methods can compromise detection accuracy, as adding noise reduces model performance. Liu et al. [13] proposed a blockchain-enabled FL system for intrusion detection in vehicular edge computing. It incorporates multi-party aggregation, trust-based incentives, and Differential Privacy to enhance accuracy and privacy. Blockchain integration introduces computational overhead, which may strain resource-constrained devices. Li et al. [14] applied a federated deep learning framework for detecting cyber threats in industrial CPS. Their CNN-GRU-based model uses Paillier encryption to ensure privacy, which significantly increases communication and processing costs. Geyer et al. [15] introduced a privacy-enhancing FL method with two phases: participant selection and Gaussian noise addition to trained models before exchanging them for a global model. This approach prevents malicious inference of participants' data. The system guarantees strong privacy, but at the cost of reduced model precision. Qu et al. [16] developed a verifiable, privacy-enhanced FL framework ensuring training integrity and confidentiality. The additional verification processes may impact scalability in large systems. Yong et al. [17] presented a privacy-preserving FL framework based on chained Secure Multi-Party Computation. This method provides strong privacy protections while maintaining performance levels comparable to baseline FL algorithms. However, Secure Multi-Party Computation is vulnerable to inference attacks as adversaries can deduce sensitive information from encrypted communications.

Existing research has not considered heterogeneous data and has mostly focused on evaluating the approaches over a single type of dataset. Evaluating the impact of different types of data on FL-based ID is important because data heterogeneity significantly influences model performance. IoT devices and edge networks often generate different data, which can lead to challenges in model convergence, accuracy, and generalisation. It is important to understand how different data types, such as attack patterns, system configurations, and network

behaviours. These features affect the performance of FL-based IDS while ensuring privacy with minimal communication overheads.

III. Proposed Framework and Threat Model

In this section, we describe our proposed framework and the associated threat model.

A. System Architecture

The centralised systems may have privacy concerns as raw data from the participating or collaborating devices is transferred to the centralised system for processing and applying machine learning. The centralised system is not only a single point of attack or failure, but the centralised system could also misuse the data without the user's consent. FL mitigates security and privacy risks by processing data on the client's device directly. FL-based systems facilitate collaboration between edge devices or end-users without using their private data. In FL setup, there can be several data owners or collaborators such as $\{F_1, \ldots, F_N\}$, which process the data $\{D_1, \ldots, D_N\}$ and train the ML model $\{M_1, \ldots, M_N\}$ locally at the each device or edge. The model M_n is then exchanged to a centralised system for an aggregate model update. The centralised system aggregates them and updates devices with the aggregated model

In traditional FL, all participating devices contribute equally to the global model. However, in a real-world scenario, collaborators are not contributing equally to the global model. Some devices may have trained their local model on several data points, and others may have trained their model over a limited or noisy dataset. This scenario should be considered while aggregating the local models. Therefore, We used Weighted FL to minimise the impact of heterogeneous collaboration. By considering weights assigned to different edge devices or collaborators, Weighted FL ensures that the global model has incorporated the data from trusted sources with high weight and assigns a small weight to malicious collaborators. The assigned weights reflect the importance of each device's contribution and ensure that the global model is influenced more by high-quality, trustworthy, and relevant updates. The proposed Intrusion Detection System combines the privacy properties of FL and the weighted aggregation by assigning weights to collaborating devices. The system comprises two primary components: the server and the edge node, or the collaborators. The server acts as a central entity responsible for aggregating models contributed by edge nodes, assigning weight to the collaborators and coordinating the updated model with the collaborators. The edge nodes are distributed and train local models on their respective local raw data. The system architecture is represented in figure 1, which mainly consists of the collaborators (edge devices) and the centralised aggregation system. The operations are mentioned below.

1) The trusted centralised system is the major component of proposed FL-based IDS systems. The system aggregates local models reported by the edge devices and computes the global model. The weighted averaging method has been used for aggregating the scores from the collaborators.
2) The edge device is responsible for three major functions: 1) the processing of local data and applying ML training on the local data, 2) the exchange of the ML model to the centralised system, and 3) updating the local model after receiving an update from the centralised system. The edge device can use any Machine learning or AI system to train the model. In this paper, we used the Neural Networks algorithm to train the local model. The whole process ensures privacy as local models are trained independently at the local device, as raw data does not leave the edge or device.
3) Each edge device shares the computed ML model with the centralised system for global model aggregation. The raw data remains securely withheld on the edge device.
4) The centralised system receives updates from the collaborating devices and computes the global model. We used weighted averages for the global model as we have assigned different weights to different collaborators. The weights can be assigned based on the reputation and trustworthiness of collaborators. This means that devices or collaborators with higher scores are given greater influence in shaping the global model. This trust-aware weighting helps mitigate the impact of potentially malicious or low-quality contributors. The trust scores are computed by estimating the deviation of the local model from the global model parameters. Local models that significantly deviate from the expected output may indicate malicious or adversarial behaviours and would receive lower trust scores.
5) The updated global model is redistributed to the edge device or collaborators, who then apply it to their local data for the classification and decision. This iterative process continues until the model achieves the desired accuracy or convergence.

The weighted averaging mechanism can enable devices with

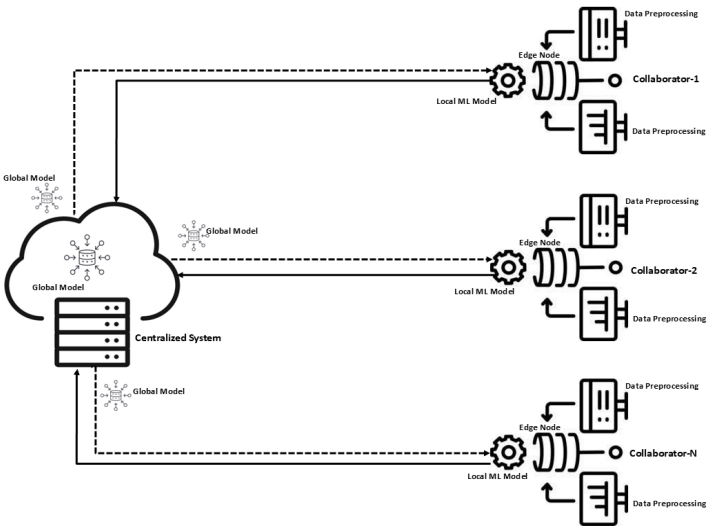


Fig. 1: FL System Architecture with Edge Nodes

less relevant data or malicious collaborators to negatively impact the performance of the global model. This approach not only improves the accuracy and reliability of the intrusion detection process but also ensures that the system achieves reliable performance even under evolving threat landscapes. The process ensures privacy as model updates cannot be used to infer the behaviour of individual collaborators.

B. Threat Model

In this paper, we used an honest-but-curious (semi-honest) threat model where we assume that the collaborating entities and the central server execute the protocol operations correctly but may try to infer sensitive information from the data or the model update. We assume that collaborating clients are not providing false results regarding model updates and that the centralised server operates honestly. We assume participants are not malicious and do not deviate from the protocol or actively disrupt the system.

IV. Experimental Setup and Performance Metrics

A. Experimental Setup

We evaluate the performance of FL-enabled IDS using widely used datasets CICIDS2017, UNSW-NB15, and KDD Cup 99 datasets. After preprocessing and cleaning, each dataset is divided into subsets to simulate multiple clients. For example, the dataset can be split by attack types across clients, and each client receives a fixed number of data points. In our experiments, we distributed the dataset across 10 clients, each receiving 10% of the specified traffic type. The splitting process has been maintained carefully so that each node contains the same class distribution across all subsets. This step is crucial to prevent data skew and biased training on some nodes. Each subset of data is assigned to an individual FL node. The FL training process involves simulating collaboration among 10 nodes for 10 iterations. Each node computes updates based on the local training and sends the model to a centralised server. At the central server, the received updates from all nodes are aggregated using Weighted Federated Averaging. The aggregated updates are sent back to the participating nodes, where edge nodes can apply them to their local data to evaluate the performance.

B. Evaluation Metrics

We used the following metrics to evaluate the performance of the system: **Accuracy:** Accuracy measures the proportion of correctly classified samples (both normal and attack) out of the total samples. A higher accuracy value indicates better overall model performance. **False Positive Rate (FPR):** refers to the proportion of benign activities incorrectly classified as malicious by an Intrusion Detection System (IDS). It is a critical metric in evaluating the performance of IDS, as a high FPR can lead to unnecessary alerts, wasted resources, and decreased system reliability. **True Positive Rate (TPR):** measures the proportion of actual intrusions correctly identified by the system as an intrusion. A high TPR ensures that most attacks are successfully detected, reducing the risk of undetected threats in the system. **Iteration:** In FL, iteration is a key performance measure which helps to understand how quickly the system is converging. In this paper, an iteration refers to a single round of communication and model updates between the central server and the participating devices.

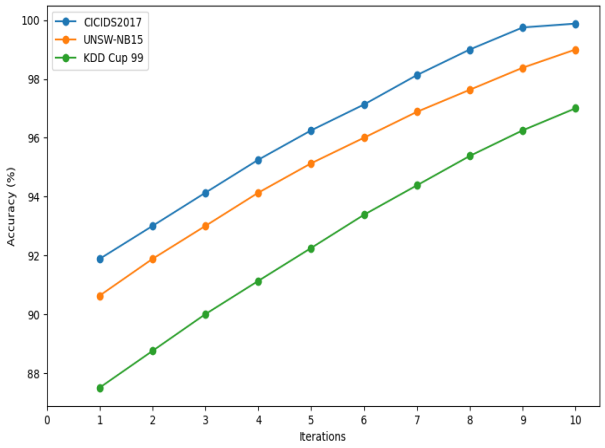


Fig. 2: Accuracy for FL under different iterations and 10 nodes

V. Performance Results

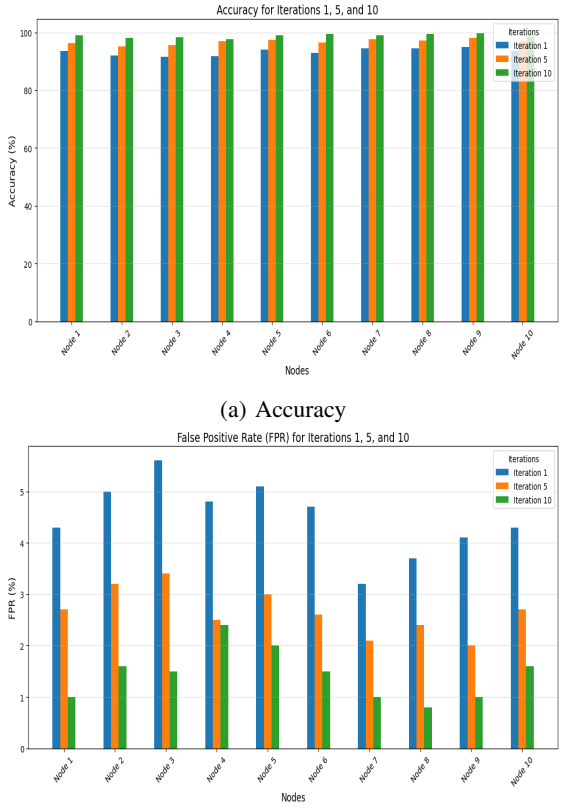
In this section, we analyse the results for different performance metrics.

A. Accuracy over Iterations

Figure 2 shows the performance of an FL model under three datasets: CICIDS2017, UNSW-NB15, and KDD Cup 99 for ten iterations and ten nodes (collaborators). In this setup, each node uses the same type of dataset one by one. The accuracy increases with the number of iterations, which shows that FL performs well over time. The CICIDS2017 dataset shows the highest accuracy throughout the iterations, reaching close to 98% by the 10th iteration. The slow improvement in accuracy shows that the global model benefits significantly from early iterations. UNSW-NB15 achieves an accuracy of 97% by the final iteration. The KDD Cup 99 dataset has the lowest accuracy, starting at around 88% and steadily reaching approximately 96% by the 10th iteration. The class imbalance in this dataset may limit its ability to leverage the full potential of the neural network models at the edge device.

The results show that the accuracy improves as the edge device collaborates with the local model through the collaboration cycle. The improved accuracy has been observed in the first few iterations (e.g., 1–4), because of the initial aggregation of local models. However, the accuracy increases very slowly during the later collaboration cycles (e.g., from 8–10 iterations). This shows that the model converges very slowly towards the model's optimal performance. This iterative collaboration shows that FL achieves higher accuracy as the local models have been trained over heterogeneous and diverse data sources.

It is important to analyse the impact of using multiple datasets in an FL setup. This would allow us to evaluate



(a) Accuracy

(b) False Positive Rate

Fig. 3: Performance evaluation under heterogeneous datasets

the behaviour of the detection model under diverse data. In this experiment, we assigned different datasets to specific groups of nodes to simulate heterogeneity. Specifically, the CICIDS2017 dataset was equally divided across four nodes, the UNSW-NB15 dataset was distributed across three nodes, and the KDD Cup 99 dataset was distributed across three nodes. This approach enabled us to evaluate the combined impact of these datasets on the accuracy and FPR across multiple iterations. Figure 3 shows the accuracy and FPR for all nodes over different iterations. It can be seen that the accuracy is increased as compared to when the model is trained on the same datasets and achieves higher accuracy as collaboration goes on. This reflects improved performance as compared to a scenario where the same dataset is distributed across all nodes. The use of diverse datasets significantly improves detection accuracy and reduces the FPR in fewer iterations. This shows that combining datasets with different traffic and attack characteristics enriches the global model's learning process. This diversity can sometimes pose challenges for model convergence, particularly in cases where attack patterns are limited or differ significantly across collaborating devices.

B. Accuracy Under Different ML Models

The objective of this experiment is to evaluate the effectiveness of an ensemble-based FL approach for intrusion detection using CICIDS2017, UNSW-NB15, and KDD Cup 99. The goal

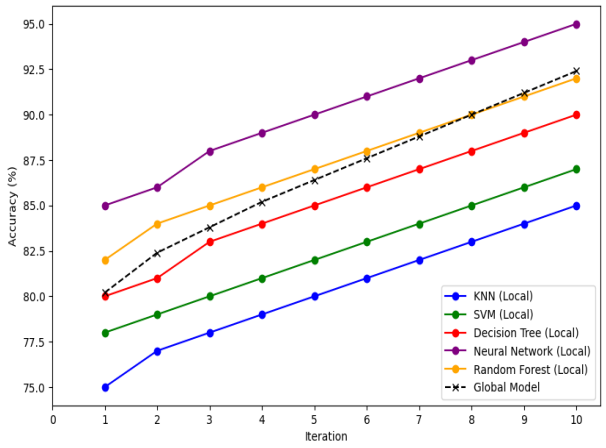


Fig. 4: Accuracy of FL systems under different Machine learning Models at Client Side.

is to simulate real-world scenarios where the global model is supplied with models trained on diverse traffic patterns and attack types. The experimental setup involves 10 edge nodes having different ML models. We used two nodes each for specific ML models, that is KNN, SVM, Decision Tree, Neural Network, and Random Forest. The experiment focuses on training localised models on subsets of these datasets at the edge nodes and aggregating them to create a global model at the centralised system. Furthermore, test data is also created by merging samples from all datasets to ensure diverse evaluation. The edge nodes retain subsets representative of their local data. After local training, the model is reported to a centralised system which applies the model to its testing data for inclusion in the global model. The global model is then sent to the edge device to apply it to the local data. The process continues until the convergence; however, we have used 10 iterations in this experiment.

Figure 4 shows accuracy at the edge nodes considering the respective ML model and the global model which aggregates these local models to improve overall accuracy. The results show that the accuracy improves over iterations and collaboration. For example, SVM begins with a baseline accuracy of (77.5%) and achieves an accuracy close to 82% after 10 iterations. Its performance improvement is consistent but slower compared to more complex models. Similarly, the neural network achieves an accuracy of 87.5% during the first iteration and manages to achieve an accuracy of around 95% by the 10th iteration. This suggests that neural networks are highly adaptable to FL frameworks due to their capacity to model complex relationships in the data.

The global model presents the aggregated result of all local models' updates, following an ensemble learning approach. Over iterations, the global model incorporates diverse insights from local models, reducing both false positives and false negatives, as seen in the steadily increasing accuracy. Its performance lies between the best-performing local models (neural network and random forest). This suggests that the global aggregation effectively combines the strengths of indi-

TABLE I: Evaluation Metrics for Intrusion Detection Across Different Numbers of Nodes

Metric	Dataset	2 Nodes	4 Nodes	6 Nodes	8 Nodes	10 Nodes
Accuracy	CICIDS2017	0.98	0.98	0.99	0.99	0.99
	UNSW-NB15	0.93	0.94	0.95	0.95	0.96
	KDD Cup 99	0.91	0.93	0.94	0.94	0.95
False Positive Rate	CICIDS2017	0.04	0.03	0.03	0.02	0.02
	UNSW-NB15	0.07	0.06	0.05	0.04	0.04
	KDD Cup 99	0.10	0.09	0.08	0.07	0.06
True Positive Rate	CICIDS2017	0.97	0.98	0.99	0.99	0.99
	UNSW-NB15	0.91	0.92	0.93	0.94	0.94
	KDD Cup 99	0.91	0.92	0.93	0.94	0.95

vidual models. It can be seen that the global model consistently outperforms local models, which are not achieving higher accuracy, as these local models are not contributing with good strength during the model aggregation. The accuracy of the global model follows the same pattern as the models which perform well at the local level, for example, it follows the random forest model, especially towards higher iterations. This indicates that random forest contributes significantly to global aggregation. It can also be noted that the global model also benefits from the diversity of insights provided by other models, especially the neural network.

C. Performance with Number of Collaborators

FL facilitates collaboration with the exchange of ML models among edge devices. The performance metrics, such as accuracy, FPR, and TPR are affected by the number of participating edge devices as shown in Table I. The TPR measures the number of samples correctly identified as malicious to the total number of malicious samples. The TPR would increase with the number of participating collaborators. For instance, in the CICIDS2017 dataset, TPR increased from 0.81 with only one collaborator to 0.96 when the number of collaborators reached 10. This increase in TPR reflects that collaboration among multiple collaborators considers diverse traffic patterns with the increased number of collaborators. In this setup, each collaborator considers their own traffic patterns and contributes towards creating a global model. Similar behaviour has been observed in the UNSW-NB15 and KDD Cup 99 datasets, where TPR increases from 0.75 to 0.93 and 0.70 to 0.93, respectively with the number of collaborators. A small FPR reflects that the system can classify normal traffic as normal with a higher percentage. The results show that FPR decreases as the number of collaborators increases. For example, in the CICIDS2017 dataset, FPR decreases from 0.18 with one node to 0.04 with ten nodes. By aggregating data from multiple collaborators, FL reduces the risk of misclassifying normal instances as attacks. Similar patterns have been seen in the UNSW-NB15 and KDD Cup 99 datasets, where FPR reduces from 0.23 to 0.06 and 0.28 to 0.07, respectively. These results highlight the ability of FL to achieve a more balanced representation of the data when more collaborators contribute to the training process. Accuracy also improves with the number of collaborators. For the CICIDS2017 dataset, accuracy increases from 0.84 with one client to 0.97 with ten clients. The same behaviour has been observed in the UNSW-NB15 and KDD Cup 99 datasets, where accuracy improves from 0.79 to 0.95 and 0.74 to 0.94, respectively. This steady increase highlights the scalability of FL as more collaborators join the collaborative process. As the number of collaborators increases, the global model benefits from greater data heterogeneity, which improves its ability to generalise across diverse scenarios.

VI. Conclusion

Traditional Intrusion Detection Systems primarily rely on centralised architectures that aggregate and process sensitive data from distributed devices. These systems could provide acceptable accuracy and effective detection but pose a serious threat to data privacy. These privacy issues highlighted the need for a model where data is processed at the device level and collaboration is achieved through exchanging the learned information, for example ML model. FL decentralises the training process by enabling local devices to train the model by using local data and then transfer the learned model to the centralised system for model aggregation and update. This decentralised approach enhances data privacy while maintaining model performance. This paper presented the FL-based IDS system for IoT networks, which enables data to be processed at the edge nodes under an honest but curious threat model. We evaluated the data for three datasets and different evaluation metrics. We used NN ML at the client level to learn the local Model and weight averaging of the centralised server while assigning different weights to collaborators. The FL approach achieves high accuracy across various metrics, comparable to traditional centralised systems. For example, CICIDS2017 achieves an accuracy of 98% after 10 iterations, with significant improvements in TPR and a small FPR. Similarly, UNSW-NB15 and KDD Cup 99 exhibit steady improvements in detection metrics, highlighting FL's ability to classify data correctly. As a part of future work, we are looking for a malicious model where the participating node might poison the data in two aspects: poisoning the model after learning, and poisoning the dataset used for model learning.

References

[1] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou, "Understanding the mirai botnet," in *Proceedings of the 26th USENIX Conference on Security Symposium*, ser. SEC'17. USA: USENIX Association, 2017, p. 1093–1110.

[2] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in internet of things," *Journal of Network and Computer Applications*, vol. 84, pp. 25–37, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804517300802

[3] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, "Network intrusion detection for iot security based on learning techniques," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2671–2701, 2019.
[4] D. Midi, A. Rullo, A. Mudgerikar, and E. Bertino, "Kalis — a system for knowledge-driven adaptable intrusion detection for the internet of things," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 656–666.
[5] E. Hodo, X. Bellekens, A. Hamilton, P.-L. Dubouilh, E. Iorkyase, C. Tachtatzis, and R. Atkinson, "Threat analysis of iot networks using artificial neural network intrusion detection system," in *2016 International Symposium on Networks, Computers and Communications (ISNCC)*, 2016, pp. 1–6.
[6] N. Moustafa, B. Turnbull, and K.-K. R. Choo, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4815–4830, 2019.
[7] E. J. Cho, J. H. Kim, and C. S. Hong, "Attack model and detection scheme for botnet on 6lowpan," in *Proceedings of the 12th Asia-Pacific Network Operations and Management Conference on Management Enabling the Future Internet for Changing Business and New Computing Services*, ser. APNOMS'09. Berlin, Heidelberg: Springer-Verlag, 2009, p. 515–518.
[8] H. Zhang, J. Ye, W. Huang, X. Liu, and J. Gu, "Survey of federated learning in intrusion detection," *Journal of Parallel and Distributed Computing*, vol. 195, p. 104976, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0743731524001400
[9] A. Khraisat, A. Alazab, S. Singh, T. Jan, and A. Jr. Gomez, "Survey on federated learning for intrusion detection system: Concept, architectures, aggregation strategies, challenges, and future directions," *ACM Comput. Surv.*, vol. 57, no. 1, Oct. 2024. [Online]. Available: https://doi.org/10.1145/3687124
[10] H. Hafi, B. Brik, P. A. Frangoudis, A. Ksentini, and M. Bagaa, "Split federated learning for 6g enabled-networks: Requirements, challenges, and future directions," *IEEE Access*, vol. 12, pp. 9890–9930, 2024.
[11] Y. Zhao, J. Zhao, M. Yang, T. Wang, N. Wang, L. Lyu, D. Niyato, and K.-Y. Lam, "Local differential privacy-based federated learning for internet of things," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8836–8853, 2021.
[12] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. Vincent Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020.
[13] H. Liu, S. Zhang, P. Zhang, X. Zhou, X. Shao, G. Pu, and Y. Zhang, "Blockchain and federated learning for collaborative intrusion detection in vehicular edge computing," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 6073–6084, 2021.
[14] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "Deepfed: Federated deep learning for intrusion detection in industrial cyber–physical systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5615–5624, 2021.
[15] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," 2018. [Online]. Available: https://arxiv.org/abs/1712.07557
[16] Y. Qu, S. R. Pokhrel, S. Garg, L. Gao, and Y. Xiang, "A blockchained federated learning framework for cognitive computing in industry 4.0 networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2964–2973, 2021.
[17] Y. Li, Y. Zhou, A. Jolfaei, D. Yu, G. Xu, and X. Zheng, "Privacy-preserving federated learning framework based on chained secure multiparty computing," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6178–6186, 2021.

Muhammad Ajmal Azad is an Associate Professor of Cyber Security in the Department of Computer Science and Digital Technologies at Birmingham City University. His research interests broadly cover the areas of network security and privacy. In the past few years, he designed systems and methods for securing telecommunication users from telemarketers, robo-callers, scammers, and spammers using behavioural modelling and social network analysis. He has also explored ways to protect the privacy of telephone users in a collaborative environment. Currently, he is involved in projects like Identity spoofing, secure reputation systems, collaborative security, security of VoIP and Next-generation Network security.

Ali Kashif Bashir received the Ph.D. degree in computer science and engineering from Korea University, Seoul, South Korea, in 2012. He is a professor of networks and cybersecurity at Manchester Metropolitan University, Manchester, U.K. In the past, he held appointments with Osaka University, Japan; Nara National College of Technology, Japan; the National Fusion Research Institute, South Korea; Southern Power Company Ltd., South Korea, and the Seoul Metropolitan Government, South Korea. His research interests include: cloud computing, NFV/SDN, network virtualization, network security, IoT, computer networks, RFID, sensor networks, wireless networks, and distributed computing. Dr. Bashir is the Editor-in-Chief of the IEEE Internet Technology Policy Newsletter and the IEEE Future Directions Newsletter. He is an Editorial Board Member of journals, such as IEEE Access, the Journal of Sensor Networks, and Data Communications.

R. Muhammad Atif Azad is a Professor of Artificial Intelligence at Birmingham City University. He specialises in Nature-Inspired Optimisation, particularly Grammatical Evolution, Genetic Programming and Genetic Algorithms and various branches of Artificial Intelligence, including Adversarial Attacks on Machine Learning and Natural Language Processing. He is among the top 100 authors on Genetic Programming. His work on automatic parallel code generation won the prestigious ACM SIGEVO Silver Humies Award at the A-ranked GECCO, in 2015, with a cash prize of US$ 3,500. He has also received the Best Reviewer Award at the European Conference on Genetic Programming (EuroGP) 2015. He has led multiple major projects funded by the Office for Students, UK to further the National Artificial Intelligence Strategy, UK Digital Strategy, and BCU's 2025 strategy for creating an inclusive digital economy.

Syed Attique Shah is working as a Senior Lecturer at the School of Computing and Digital Technology, Birmingham City University (BCU), UK, and also holds the role of programme/course leader for the MSc Advanced Computer Network at BCU. He previously worked as a Lecturer/Assistant Professor at the Institute of Computer Science, University of Tartu, Estonia, and also served as an Associate Professor and Chairperson for the Department of Computer Science at BUITEMS, Pakistan. He received his PhD from the Informatics Institute, Istanbul Technical University, Turkey. He has published more than 25 research papers in reputable Q1 journals with an accumulated impact factor of more than 150. He was also mentioned in the top 2% of scientists worldwide by Stanford University and Elsevier. His area of research includes: machine learning, big data analytics, the Internet of Things (IoT), digital twining, network security and image recognition.