Federated Learning for Privacy-Preserving Collaborative Intrusion Detection

Ayaan M. Siddiqui[1], Charu N. Singhal[2], Devendra K. Tomar[3]

ABSTRACT

An intrusion-detection model learns best from large and varied data, yet the organisations that hold such data are reluctant, and often legally unable, to share their network records, which are sensitive and may reveal their vulnerabilities, so that each organisation is left to train on its own limited data. Federated learning offers a way out of this dilemma, allowing many organisations to train a shared model collaboratively without any of them disclosing its raw data. This study applies federated learning to collaborative intrusion detection and compares it with the alternatives of centralized training, which requires data to be pooled, and local-only training, in which each organisation learns alone. Several participants, each holding its own intrusion data, repeatedly trained a shared detection model locally and combined their improvements through a central coordinator, exchanging only model updates and never the data itself, and the accuracy of the resulting model was measured over successive rounds and set against the centralized and local-only baselines. The federated model improved steadily as the rounds proceeded, approaching the accuracy of the centralized model that had access to all the pooled data and substantially exceeding the accuracy that any participant achieved by training on its own data alone, while no participant ever revealed its raw records. The study shows that federated learning enables organisations to build an effective shared intrusion detector without sacrificing the privacy of their data, reconciling the need for collaboration in cyber defence with the requirements of digital trust and confidentiality.

KEYWORDS: *Federated learning, Intrusion detection, Privacy preservation, Collaborative defence, Digital trust, Distributed learning, Data confidentiality, Model aggregation, Cyber defence*

INTRODUCTION

An effective intrusion-detection model must learn the many forms that normal and malicious network activity can take, and it learns this best from large and varied data drawn from many sources. The organisations that hold such data, however, are generally unwilling to share it. Network records are sensitive: they may contain confidential information, reveal the structure and weaknesses of an organisation's systems, and expose it to legal and reputational risk if disclosed, and regulations often forbid their transfer. As a result, each organisation tends to train its detection model on its own data alone, which is limited in volume and variety and yields a weaker model than the pooled data of many organisations would.

This creates a dilemma for cyber defence. Collaboration would benefit all, because attacks seen by one organisation may threaten others and a model trained on the combined experience of many would detect more threats, yet the very sensitivity of the data that makes collaboration valuable also prevents it. Federated learning offers a resolution. In federated learning, the participants do not pool their data; instead, each trains a shared model on its own data locally and sends only the resulting improvements to the model, not the data, to a coordinator that combines the improvements from all participants into a better shared model, which is then returned for further training. The data never leaves the organisation that owns it.

This paper applies federated learning to collaborative intrusion detection and compares it with centralized training, which requires the data to be pooled, and with local-only training, in which each organisation learns alone. Several participants, each holding its own data, train a shared model collaboratively by exchanging only model updates, and the accuracy of the resulting model is measured over successive rounds and against the two baselines. The aim is to determine whether federated learning can build an effective shared detector without disclosing any participant's data, and so reconcile collaboration in cyber defence with the requirements of digital trust and confidentiality.

LITERATURE REVIEW

Federated learning was introduced as a means of training a shared model across many devices or organisations without centralising their data, and it has since been studied widely, including for security applications. Its principle is that learning can be moved to the data rather than the data to the learning: each participant computes improvements to a shared model from its own data, and these improvements, rather than the data, are combined. This preserves the privacy of the data while still allowing the model to benefit from the experience of all participants.

Federated Learning for Security

The literature identifies the main features of federated learning relevant to collaborative defence:

- The raw data never leaves the participant that owns it, so confidentiality and regulatory constraints are respected.
- Only model updates are exchanged and combined, typically by averaging, into an improved shared model.
- The shared model benefits from the combined data of all participants, approaching the performance achievable by pooling the data.
- Challenges include the uneven and non-identical distribution of data across participants and the communication cost of many rounds of exchange.

For intrusion detection in particular, federated learning is attractive because it allows organisations to pool their experience of attacks without revealing their sensitive network data, which they are otherwise unable or unwilling to share.

Studies of federated learning for intrusion and threat detection report that a federated model can approach the accuracy of a model trained on the pooled data and substantially exceed that of models trained by each participant alone, while keeping the data private, and that the chief difficulties are the heterogeneity of data across participants and the cost of communication. The literature also notes that the exchange of model updates, though far less revealing than

the data, requires care to prevent indirect leakage, and that additional protections can be applied to the updates where the threat warrants.

RESEARCH GAP

Although federated learning is established in principle, focused studies that apply it to intrusion detection and compare it directly with both the centralized and the local-only alternatives on the same data, showing how the federated model improves over successive rounds and where it stands relative to those baselines, remain valuable for demonstrating its benefit for collaborative defence. There is therefore value in a study that quantifies the accuracy of federated intrusion detection against pooled and isolated training while making explicit that no raw data is shared.

OBJECTIVES

The objectives of this study are:

- To apply federated learning to the collaborative training of a shared intrusion-detection model across several participants.
- To train the shared model by exchanging only model updates, never the raw data.
- To measure the accuracy of the federated model over successive rounds of training.
- To compare the federated model with centralized training on pooled data and with local-only training by each participant alone.

METHODOLOGY

Federated Setup

Several participants were modelled, each holding its own set of intrusion data representing the activity it observed, the data differing in distribution between participants. A shared detection model was trained by federated learning: in each round, every participant trained the current shared model on its own data and sent the resulting update to a central coordinator, which combined the updates from all participants by averaging into an improved shared

model and returned it for the next round. The raw data of each participant never left it. The setup is summarised in Table 1.

Table 1: Federated learning setup for collaborative intrusion detection

Element	Description
Participants	Several, each with its own intrusion data
Data distribution	Non-identical across participants
Exchanged	Model updates only, never raw data
Aggregation	Averaging of updates by the coordinator
Training	Repeated rounds of local training and aggregation

Baselines and Evaluation

The federated model was compared with two baselines: a centralized model trained on the pooled data of all participants, representing the accuracy achievable if the data could be shared, and local-only models, each trained by a single participant on its own data alone, representing the accuracy achievable without collaboration. The detection accuracy of the federated model was measured after each round and compared with the centralized and local-only baselines, and the privacy of each approach, in terms of whether raw data was disclosed, was noted.

RESULTS AND FINDINGS

The federated model improved steadily as the rounds of training proceeded. As shown in Figure 1, its detection accuracy rose from a low value at the outset, when the shared model had learned little, and increased with each round as the participants' improvements were

combined, until it levelled off near the accuracy of the centralized model. Throughout this process no participant disclosed its raw data; only updates to the model were exchanged.

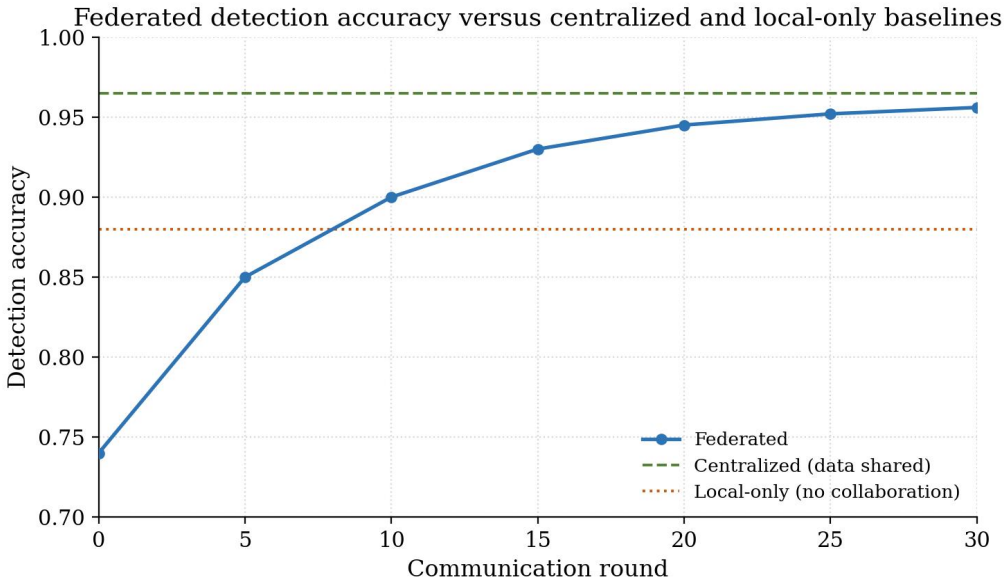

Figure 1: Detection accuracy of the federated model over successive rounds, against the centralized and local-only baselines

At convergence the federated model approached the accuracy of the centralized model that had access to all the pooled data, falling only slightly short of it, and it substantially exceeded the accuracy that any participant achieved by training on its own data alone. The small shortfall relative to the centralized model is the modest price of not pooling the data, attributable to the uneven distribution of data across participants, and is far outweighed by the large gain over local-only training. The comparison of the three approaches is summarised in Table 2.

Table 2: Comparison of the three training approaches

Approach	Accuracy	Raw data shared	Collaboration
Local-only	0.880	None	None

Approach	Accuracy	Raw data shared	Collaboration
Federated	0.956	None	Yes
Centralized (pooled)	0.965	All	Yes

DISCUSSION

The results show that federated learning resolves the dilemma between collaboration and confidentiality in intrusion detection. The federated model, trained on the combined experience of all participants without any of them disclosing its data, approached the accuracy of the centralized model that required the data to be pooled and greatly exceeded the accuracy that each participant could achieve alone. Organisations can therefore obtain nearly the full benefit of collaboration, a detector informed by the attacks seen across all participants, while retaining the privacy of their sensitive network data, which is what makes collaboration possible where it would otherwise be forbidden or refused.

The steady improvement of the federated model over the rounds illustrates how the approach works. At the start the shared model is uninformed, but each round combines what every participant has learned from its own data, so that the model accumulates the experience of all without ever seeing their data directly. The small shortfall relative to the centralized model reflects the cost of learning from data that is distributed and unevenly shared between participants rather than gathered in one place, a known effect of data heterogeneity in federated learning, but this shortfall was minor and the federated model remained far superior to any local-only model. The large gain over local-only training is the essential result, since local-only training is the realistic alternative when data cannot be shared.

The practical and trust implications are significant. Federated learning allows organisations that compete or are bound by confidentiality and regulation to nonetheless cooperate in defence, pooling their knowledge of threats without pooling their data, which strengthens the security of all while respecting the privacy of each. This reconciles the collaborative

imperative of cyber defence, that a threat to one is a threat to many, with the requirements of digital trust and data protection. Some caveats remain: the exchange of model updates, while far less revealing than the data, should be protected against indirect leakage where the threat warrants, the heterogeneity of data across participants limits how closely the federated model can match the centralized one, and the approach incurs the communication cost of many rounds of exchange. Within these limits, federated learning is a powerful means of collaborative, privacy-preserving intrusion detection.

LIMITATIONS

The following limitations apply to this study:

- The participants and their data were modelled, and real deployments would involve greater scale and more varied data distributions.
- The exchange of model updates was assumed not to leak sensitive information, whereas in adversarial settings additional protection may be required.
- The heterogeneity of data across participants affects the result, and only a representative degree of it was considered.
- The communication cost of repeated rounds was noted but not optimised, and it may be significant at scale.

FUTURE SCOPE

Future work could extend the present study as follows:

- Application of additional protections to the exchanged updates to guard against indirect information leakage.
- Techniques to cope with strong heterogeneity of data across participants and to narrow the gap to centralized training.
- Reduction of the communication cost through more efficient update exchange and aggregation.
- Evaluation at larger scale with many real participants and diverse threats.
- Resistance to participants that behave dishonestly or attempt to corrupt the shared model.

CONCLUSION

Federated learning was applied to collaborative intrusion detection and compared with centralized training on pooled data and with local-only training by each participant alone. Several participants trained a shared detection model by exchanging only model updates, never their raw data, and the federated model improved steadily over successive rounds, approaching the accuracy of the centralized model with access to all the pooled data and substantially exceeding the accuracy that any participant achieved alone, while no raw data was ever disclosed. The study shows that federated learning enables organisations to build an effective shared intrusion detector without sacrificing the privacy of their data, reconciling the collaborative imperative of cyber defence with the requirements of digital trust and confidentiality, subject to protecting the exchanged updates where warranted, to the effect of data heterogeneity on the small shortfall relative to centralized training, and to the communication cost of repeated rounds.

REFERENCES

1. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Aguera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 1273–1282.
2. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 14(1–2), 1–210.
3. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2), 12.
4. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, 1175–1191.
5. Nguyen, T. D., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N., & Sadeghi, A. R. (2019). DIoT: A federated self-learning anomaly detection system for IoT. *Proceedings of the IEEE International Conference on Distributed Computing Systems*, 756–767.

6. Mothukuri, V., Khare, P., Parizi, R. M., Pouriyeh, S., Dehghantanha, A., & Srivastava, G. (2022). Federated-learning-based anomaly detection for IoT security attacks. *IEEE Internet of Things Journal*, 9(4), 2545–2554.
7. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), 50–60.
8. Rey, V., Sanchez Sanchez, P. M., Huertas Celdran, A., & Bovet, G. (2022). Federated learning for malware detection in IoT devices. *Computer Networks*, 204, 108693.
9. Liu, Y., Kumar, N., Xiong, Z., Lim, W. Y. B., Kang, J., & Niyato, D. (2020). Communication-efficient federated learning for anomaly detection in industrial Internet of Things. *Proceedings of the IEEE Global Communications Conference*, 1–6.
10. Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y. C., Yang, Q., Niyato, D., & Miao, C. (2020). Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(3), 2031–2063.

***Author for Correspondence**
Ayaan M. Siddiqui
E-mail: ayaan.siddiqui66@gmail.com

[1]Associate Professor, Department of Computer Science Engineering, Mahatma Gandhi Mission's College Of Engineering & Technology, Noida , NCR
[2]Assistant Professor, Department of Computer Science and Engineering, Mahatma Gandhi Mission's College Of Engineering & Technology, Noida , NCR

[3]Research Scholar, Department of Information Technology, Mahatma Gandhi Mission's College Of Engineering & Technology, Noida , NCR

Received Date: May 24, 2026
Accepted Date: June 08, 2026
Published Date: June 20, 2026

Citation: Ayaan M. Siddiqui, Charu N. Singhal, Devendra K. Tomar. Federated Learning for Privacy-Preserving Collaborative Intrusion Detection. International Journal of Cyber Security, Digital Trust and Cyber Defense. 2026; 2(1): 33-42p.