© APR 2026 | IRE Journals | Volume 9 Issue 10 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I10-1717536

Federated Learning Paradigms for Privacy-Preserving Multi-Organizational Threat Intelligence Sharing

MARCELO ARAUJO

Abstract- ***Cyber threat intelligence sharing is widely recognized as a strategic component for improving early detection of malicious campaigns, correlation of indicators of compromise, and coordinated incident response. Despite this potential, direct exchange of operational data among institutions remains constrained by regulatory, contractual, competitive, and technical barriers, especially when network telemetry, authentication logs, endpoint events, and sensitive artifacts are involved. In this context, federated learning has been investigated as an approach capable of enabling collaborative training without centralizing raw data. This article discusses the main federated learning paradigms applied to multi-organizational cyber threat intelligence sharing, with emphasis on privacy preservation, robustness against adversarial manipulation, statistical heterogeneity across participants, and scalability limitations. It also examines complementary techniques such as secure aggregation, differential privacy, homomorphic encryption, secure multi-party computation, and Byzantine-robust mechanisms. Recent literature suggests that federated learning can improve the generalization capability of detection models when compared with strictly local approaches, although its practical adoption still depends on more mature solutions for inter-organizational trust, semantic interoperability, and technical governance.***

Keywords: Federated Learning, Cyber Threat Intelligence, Differential Privacy, Secure Aggregation, Intrusion Detection.

I. INTRODUCTION

Cyber threat intelligence increasingly depends on the ability to integrate distributed signals observed across different operational environments. Contemporary attacks rarely remain confined to a single organization, since phishing campaigns, vulnerability exploitation, modular malware, ransomware, and lateral movement often leave partial traces across multiple sectors at the same time. For this reason, collaboration among organizations has become relevant for expanding visibility, reducing informational asymmetries, and anticipating attack trends [6,10].

However, conventional threat intelligence sharing faces major constraints. In many cases, the data required to train security models include commercially, operationally, or legally sensitive information. Authentication logs, customer metadata, internal fraud indicators, traffic patterns, and incident records may contain elements that cannot be freely transferred to a central repository. This obstacle is especially significant in sectors such as finance, healthcare, telecommunications, and critical infrastructure, where confidentiality and compliance requirements are particularly strict [2,4,8].

In this setting, federated learning emerged as a technical alternative capable of supporting collaborative analytics without centralizing raw data. Instead of sending local datasets to a central server, each organization trains a model within its own infrastructure and shares only parameter updates, gradients, or derived representations, which are later aggregated into a global model [1,2]. This architectural shift does not eliminate every risk, but it reduces direct data exposure and makes multi-organizational knowledge sharing more feasible.

Federated learning and its application to CTI sharing

The classical federated learning model is organized in iterative rounds. First, a global model is distributed to participants. Then, each client performs local training using its own data and sends updates to the server, which aggregates them to produce the next global version. FedAvg remains a fundamental reference in this context because it showed that decentralized learning can be combined with reduced communication cost when compared with continuous synchronization approaches [1].

In cyber threat intelligence, the most common setting tends to be cross-silo federated learning, in which the number of participants is relatively small, but each entity operates its own infrastructure, holds larger datasets, and faces stricter governance requirements [2]. This arrangement is more consistent with business

© APR 2026 | IRE Journals | Volume 9 Issue 10 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I10-1717536

consortia, sector-specific threat-sharing centers, and cooperation among regulated institutions.

The application of federated learning to CTI nevertheless introduces specific challenges. Data collected by different organizations rarely follow the same statistical distribution. Beyond quantitative variation, there are differences in operational context, network architecture, security maturity, event typology, and labeling granularity. As a result, a hospital, a financial institution, and an energy provider are likely to observe very different patterns, producing a non-IID setting that affects global model convergence and stability [2,13,14]. Even so, recent studies show that federated models can outperform strictly local models in intrusion detection, anomalous traffic classification, and attack pattern recognition, precisely because distributed training expands the diversity of signals used in learning [6,7,11,12]. To clarify this workflow, Figure 1 summarizes how participating organizations train locally, share only protected model updates, and contribute to a global CTI model without transferring raw data.

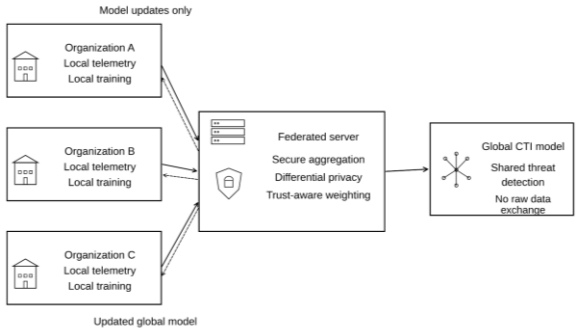

Figure 1. Simplified federated learning workflow for privacy-preserving multi-organizational cyber threat intelligence sharing

Source: Created by author.

Sarhan et al. demonstrated that a CTI sharing scheme based on federated learning can improve the performance of intrusion detection systems by enabling cooperation among organizations without full data centralization [6]. Later studies reinforced this direction by exploring decentralized cybersecurity settings, heterogeneous IoT environments, and knowledge distillation mechanisms to mitigate the effects of non-homogeneous distributions [7,11,12]. Therefore, the value of federated learning in CTI does not lie only in privacy preservation, but also in expanding statistical coverage and sensitivity to emerging patterns.

Privacy preservation and the limits of the federated model

Although federated learning is often described as a privacy-preserving technique, keeping data at the source does not guarantee complete protection. Model updates may reveal information about local training datasets, especially when adversaries exploit gradient inference, membership inference, or partial reconstruction attacks [2,4,5]. Therefore, in sensitive applications, the literature recommends combining federated learning with additional protective mechanisms.

Among these mechanisms, secure aggregation plays a central role. The protocol proposed by Bonawitz et al. became a reference because it allows the server to access only the aggregated result of client updates, without individually viewing each participant's parameter vector [3]. In practical terms, this reduces the risk of improper inspection by a central orchestrator and strengthens confidentiality among institutions that do not want to expose details of their internal security patterns.

Another widely used technique is differential privacy, which injects calibrated noise into shared updates to limit the possibility of inferring the contribution of individual records [4]. In CTI environments, this is especially relevant because rare indicators or highly specific events may, under some circumstances, reveal the occurrence of a particular incident or the existence of a sensitive asset. However, applying differential privacy requires careful balance between utility and protection. Stronger noise levels may impair the model's ability to detect minority classes, which is particularly problematic in cybersecurity, where rare events often have high analytical value [4,8,9]. Specifically, the mathematical calibration of the privacy budget dictates a strict trade-off: tightening the privacy bounds to prevent the reconstruction of individual network events inevitably degrades the model's sensitivity to zero-day exploits and highly targeted, low-frequency indicators of compromise [4, 8]

In addition, homomorphic encryption and secure multi-party computation are discussed as

complementary mechanisms for strengthening collaborative processing without exposing either raw data or individual updates [5,8,14]. These approaches enhance the formal security of the system, but they also tend to introduce computational overhead, latency, and implementation complexity. Their adoption therefore depends on the balance between contextual sensitivity, institutional operational capacity, and near-real-time performance requirements.

Camalan and Celiktas proposed a framework that combines private information retrieval, federated learning, and differential privacy for cyber threat intelligence applications, emphasizing the need for multiple protection layers in collaborative ecosystems [8]. Similarly, Pandey et al. discussed a privacy-preserving model for CTI sharing across multi-organizational platforms [9]. These contributions indicate that federated learning should be understood as part of a broader security architecture, rather than as a self-sufficient solution.

Robustness, trust, and adversarial manipulation

A critical point in applying federated learning to cyber threat intelligence is the possibility of malicious or faulty participation. Participating organizations may be compromised, misconfigured, or exposed to attacks that contaminate local training. In such situations, the aggregation server may receive manipulated updates intended to degrade the global model, alter its sensitivity to certain classes, or introduce backdoor patterns [5,10,14].

The distributed learning literature has shown that naive linear aggregation methods are vulnerable in the presence of Byzantine agents. In response, robust algorithms such as Krum, trimmed mean, and median-based aggregation were proposed to reduce the influence of extreme or anomalous updates [15,16]. In CTI environments, such methods are especially relevant because trust among participants is not always homogeneous, and the very nature of the domain assumes the presence of sophisticated adversaries. While these statistical distance-based methods effectively filter out random Byzantine faults or misconfigured nodes, they may struggle against sophisticated adversaries executing stealthy data poisoning or label-flipping attacks designed to evade standard anomaly thresholds [15, 16].

Mrabet proposed the TrustFed-CTI framework, which introduces a dynamic trust evaluation component into the aggregation process [10]. Instead of relying only on local dataset volume, the contribution of each client can be weighted by reputational indicators, historical consistency, performance stability, and signs of suspicious behavior. This approach makes the system more compatible with real cooperation scenarios, in which algorithmic governance must consider the quality, credibility, and resilience of incoming contributions.

This issue is particularly important to avoid excessive claims about federated learning performance. Although some studies report high accuracy levels on specific datasets, these results should not be generalized as universal guarantees. In cybersecurity, model effectiveness strongly depends on threat type, labeling quality, local data representativeness, and the ability to cope with temporal drift. Therefore, absolute claims regarding accuracy or robustness should be replaced by more context-dependent formulations grounded in the available evidence [6,10-12].

Scalability, heterogeneity, and interoperability

Beyond privacy and robustness, federated learning also faces relevant challenges regarding communication efficiency and interoperability. Each training round requires parameter exchange, synchronization among participants, consistency verification, and, in some cases, fault handling for missing or slow nodes. In multi-organizational environments, these difficulties are intensified by differences in infrastructure, internal security policies, data formats, and computational availability [2,13,14].

Recent literature points to several strategies for mitigating these issues, including update compression, hierarchical aggregation, partial client selection, and hybrid edge-cloud architectures [2,13]. Even so, semantic heterogeneity remains a major obstacle. Different organizations may use distinct taxonomies for security events, incompatible ontologies, and proprietary feature extraction pipelines. As a consequence, federated learning may aggregate useful statistics, but it does not automatically solve the problem of conceptual alignment across participants. To mitigate this, federated learning pipelines must be integrated with standardized CTI representation frameworks, such as STIX (Structured Threat

Information Expression) and TAXII (Trusted Automated Exchange of Intelligence Information), ensuring that localized feature extraction pipelines maintain contextual consistency before aggregation occurs [17].

This aspect requires greater specificity in technical articles, because excessively generic formulations may suggest that federated CTI sharing depends only on algorithmic infrastructure. In practice, it also requires data governance, minimum event standardization, transparent validation criteria, and audit mechanisms. Without these elements, collaboration may generate statistically aggregated models that remain operationally difficult to interpret or deploy across distinct environments [2,10,14].

CONCLUSION

Federated learning paradigms offer a technically plausible basis for multi-organizational threat intelligence sharing with stronger privacy preservation than conventional centralized models. Their main contribution lies in enabling analytical collaboration among entities that cannot, for regulatory or strategic reasons, transfer their raw data to a single infrastructure. When combined with secure aggregation, differential privacy, cryptographic mechanisms, and Byzantine-robust techniques, these paradigms can strengthen the development of more resilient collaborative models [3-5,10,15,16].

However, the practical adoption of federated learning in CTI should not be presented in a simplified manner. Important challenges remain regarding data heterogeneity, semantic interoperability, communication cost, inter-institutional governance, and defense against adversarial manipulation. For this reason, recent literature suggests that the most promising advances will emerge from hybrid, auditable, and trust-aware architectures capable of balancing privacy, analytical utility, and operational resilience [10-14]. In summary, federated learning represents a relevant infrastructure for the future of threat intelligence sharing, provided that it is treated as one component of a broader technical and organizational ecosystem rather than as a stand-alone solution.

REFERENCES

[1] McMahan HB, Moore E, Ramage D, Hampson S, Agüera y Arcas B. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. PMLR. 2017;54:1273-82.

[2] Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, et al. Advances and open problems in federated learning. Found Trends Mach Learn. 2021;14(1-2):1-210.

[3] Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, et al. Practical secure aggregation for privacy-preserving machine learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM; 2017. p. 1175-91.

[4] Dwork C, Roth A. The algorithmic foundations of differential privacy. Found Trends Theor Comput Sci. 2014;9(3-4):211-407.

[5] Nguyen T, Thai MT. Preserving privacy and security in federated learning. IEEE/ACM Trans Netw. 2024;32(1):833-43.

[6] Sarhan M, Layeghy S, Moustafa N, Portmann M. Cyber threat intelligence sharing scheme based on federated learning for network intrusion detection. J Netw Syst Manage. 2023;31:1-23.

[7] Sakhare NN. A decentralized approach to threat intelligence using federated learning in privacy-preserving cyber security. J Electr Syst. 2024;20(2):658.

[8] Camalan E, Celiktas B. Privacy-preserving cyber threat intelligence: a framework combining private information retrieval, federated learning, and differential privacy. In: 2025 10th International Conference on Computer Science and Engineering (UBMK). New York: IEEE; 2025. p. 1525-30.

[9] Pandey S, Azath H, Rahman R, Lamkuche H. Privacy-preserving model for cyber threat intelligence sharing across multi-organizational platforms. In: 2025 IEEE 14th International Conference on Communication Systems and Network Technologies (CSNT). New York: IEEE; 2025. p. 437-42.

[10] Mrabet M. TrustFed-CTI: a trust-aware federated learning framework for privacy-preserving cyber

threat intelligence sharing across distributed organizations. Future Internet. 2025;17(11):512.

[11] Timofte EM, Dimian M, Puscasu M, et al. Federated learning for cybersecurity: a privacy-preserving approach. Appl Sci. 2025;15(12):6878.

[12] Peng H, Wu C, Xiao Y. FD-IDS: federated learning with knowledge distillation for intrusion detection in non-IID IoT environments. Sensors. 2025;25(14):4309.

[13] Collins E, Wang M. Federated learning: a survey on privacy-preserving collaborative intelligence. arXiv [Preprint]. 2025:arXiv:2504.17703.

[14] Prajapati N. Federated learning for privacy-preserving cybersecurity: a review on secure threat detection. Int J Adv Res Sci Commun Technol. 2025.

[15] Blanchard P, El Mhamdi EM, Guerraoui R, Stainer J. Machine learning with adversaries: Byzantine tolerant gradient descent. In: Advances in Neural Information Processing Systems 30. Red Hook: Curran Associates; 2017. p. 119-29.

[16] Yin D, Chen Y, Kannan R, Bartlett P. Byzantine-robust distributed learning: towards optimal statistical rates. In: Proceedings of the 35th International Conference on Machine Learning. PMLR. 2018;80:5650-9.

[17] OASIS Cyber Threat Intelligence (CTI) Technical Committee. STIX Version 2.1. OASIS Standard; 2021.

[18] Available from: https://docs.oasis-open.org/cti/stix/v2.1/stix-v2.1.html.