Federated Learning for Privacy-Preserving Threat Intelligence Sharing among Organizations

Diogo Aniekwe

Senior Information Security Analyst, Digiss LLC, Nigeria

DOI: https://doi.org/10.51244/IJRSI.2026.1304000034

Received: 02 April 2026; Accepted: 08 April 2026; Published: 27 April 2026

ABSTRACT

Cyber threat intelligence empowers diverse organizations to detect and respond to new and evolving threats proactively. However, issues surrounding privacy and compliance restrictions nonetheless hinder practical implementation. This study explores the adoption of Federated Learning (FL) as a privacy-preserving alternative to traditional CTI sharing strategies. FL protects data sovereignty and conforms to privacy regulations by enabling multiple participating organizations to collaboratively train threat detection models together without disclosing sensitive information. The proposed approach promotes confidentiality while preserving threat detection accuracy by incorporating FL with mechanisms such as differential privacy and secure aggregation. This work highlights a conceptual system design for FL CTI sharing, analyzes the trade-offs between accuracy and privacy, and simulates how models can be evaluated for accuracy in a non-IID environment. The findings reveal that while privacy-preserving mechanisms result in acceptable performance degradation, personalized federated learning (FL) models enhance per-client accuracy in a multi-data setting. This study contributes a secure, flexible, and compliant approach to collaborative CTI sharing among organizations in diverse industries.

Keywords: Federated Learning, Threat Intelligence Sharing, Privacy-Preserving, Machine Learning, Non-IID Data

INTRODUCTION

Cyber Threat Intelligence refers to the collection, analysis, and dissemination of information related to existing and potential cyber threats. It encompasses information about attack vectors, vulnerabilities, threat actors, Indicators of Compromise (IOCs), and remediation techniques. The primary objective of CTI activities is to enhance the security posture of organizations by equipping them to detect, respond to, and predict cyber threats more effectively. Subject to the kind of threat data and the audience being addressed, CTI can be designated as tactical, strategic, operational, or technical.

Cyber threat intelligence sharing across organizations has evolved into an invaluable defense mechanism in today's multifaceted threat landscape. Organizations can detect and combat threats more effectively when they collaborate to exchange information about threats, rather than working independently. Shared intelligence fosters broader situational awareness, enabling rapid incident detection and response, identifies threat patterns across various sectors, and accelerates threat detection (Alaeifar et al., 2023).

Nevertheless, cyber threat intelligence sharing remains limited in practice due to several significant barriers, despite its apparent benefits. Among the primary challenges is the conflict between collaboration and data privacy. Sharing intelligence may unintentionally disclose confidential, sensitive details, including client information, vulnerabilities in internal systems, risks that can lead to reputational damage, or intellectual property data. Data protection laws, including the NDPR, GDPR, HIPAA, and others, impose strict restrictions on the exchange of certain types of data and carry legal and regulatory sanctions for defaulters (Trocoso-Pastoriza et al., 2022; Dash et al., 2022; Nigeria Data Protection Regulation, 2019).

This study investigates how Federated Learning (FL) can enhance the need for collaborative cybersecurity activities and data security requirements by implementing measures such as differential privacy and secure aggregation. Furthermore, it raises a key question: Does Federated learning offer a reliable and efficient

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

alternative to traditional CTI sharing methods? Within this context, this study contributes to the ongoing efforts to mitigate the limitations between collaborative cybersecurity and compliance needs.

BACKGROUND AND MOTIVATION

CTI sharing is crucial in helping organizations proactively build defense systems against cyber threats. Nevertheless, despite its generally accepted relevance, the real-world adoption of CTI sharing is still hindered by significant challenges. Challenges around privacy concerns, compliance implications, credibility, and a deficiency in standardized structures. To tackle these challenges, recent studies have recommended a range of approaches, including decentralized systems and standardized initiatives, as well as privacy-preserving machine learning strategies (Sukhabogia & Anusha, 2021; Abu et al., 2018).

The primary issue associated with CTI sharing is the wide range of data sources and varying levels of trust that accompany them. Abu et al. (2018) classified CTI into three main groups: internal CTI (SIEM, email, firewall logs, etc.); external CTI (CTI from commercial providers or open-source intelligence, etc.); and community-based CTI (CTI from trusted peer-to-peer channels, etc.). Internal CTIs offer higher visibility into threats, whereas external and community CTIs may require additional vetting and may have quality deficiencies. These issues underscore the necessity of CTI sharing to follow secure, verifiable, and privacy-preserving methods.

Sukhabogia and Anusha (2021) provided a selection of standards that support the machine-readable exchange of CTI data, including STIX, TAXII, OpenIOC, and IODEF. For example, STIX provides a modular format for communicating IOCs, and TAXII is a protocol for conveying STIX information. Despite their potential, these frameworks often encounter issues with acceptance due to their complexity and the need for integration into multiple environments.

Alaeifar et al. (2023) categorized CTI sharing mechanisms into peer-to-peer, peer-to-repository, and hybrid architectures. Their work underscores the importance of joint models, which support proactive defense, enhance situational awareness, and facilitate the rapid dissemination of IOCs. The researchers also identified some non-technical barriers, consisting of high implementation costs, a lack of incentives, and concerns about reputational damage. They claim that instruments such as monetary incentives, social endorsements, or more solid platforms for data governance can upgrade participation in CIT sharing.

There is a growing potential for automated threat detection, aggregation, and predictive analysis due to the rising intersection of ML in CTI. Consolidating vast volumes of raw data is a requirement for conventional machine learning models, which poses significant privacy concerns. Dash et al. (2022) described FL as a fix that lets organizations collectively train models without releasing their raw data sets. Their work demonstrates how FL adheres to anonymity and data minimization standards, making it ideal for systems that handle sensitive data, such as Fintechs.

Trocoso-Pastoriza et al. (2022) advanced on this school of thought by providing a federated intelligence sharing platform that features FL with cryptographic methods like differential privacy (DP), homomorphic encryption, and secure multi-party computation (SMC). Their methodology grants participating organizations full autonomy over their local data, while allowing them to compute consolidated threat insights. The approach confronts reliability and privacy concerns that plague centralized CTI sharing techniques despite the high degree of computation involved. Their exploration of many scenarios, including sharing MISP event statistics and training of DDoS detection models, is both practical and efficient.

EXISTING THREAT INTELLIGENCE SHARING STANDARDS

In cybersecurity, sharing threat intelligence is crucial for organizations to defend against threats effectively. MITRE developed some traditional threat intelligence sharing standards to strengthen collective defense capacities and ease the dissemination of security-related intelligence (Mavroeidis et al., 2020). Structured Threat Information Expression (**STIX**), Cyber Observable eXpression (**CybOX**), and Trusted Automated Exchange of Indicator Information (**TAXII**) are some of the extensively used standards in threat intelligence sharing. Organizations adopt these standards based on their individual needs. Similarly, the U.S. federal government

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

initiated Information Sharing and Analysis Centers **(ISACs)**. Although ISAC is not a standard per se, it equally serves as a viable mechanism for sharing cyber threat intelligence within specific industry sectors.

- **Structured Threat Information Expression (STIX)**

STIX is a standardized language for analyzing and communicating CTI. It functions as a guide for reporting threat-related information, including attack trends, Indicators of Compromise (IoCs), adversary profiles, and their tactics, techniques, and procedures (TTPs). With STIX, organizations can transmit threat data in a machine-readable format, enabling the rapid automation of analysis, response, and the exchange of CTI.

- **Trusted Automated Exchange of Indicator Information (TAXII)**

Paired with STIX, TAXII is a language protocol that facilitates the automated sharing of CTI. It illustrates the systematic and safe approach through which threat information can be exchanged between parties. TAXII guarantees that organizations can share CTI rapidly via real-time broadcast of findings through HTTP or HTTPS communication protocols. This standard is primarily used to disseminate Information About Security incidents, malicious activities, and vulnerabilities.

Some Challenges in Traditional CTI Sharing

- **Quality vs. Quantity of Shared Intelligence**

The effectiveness of threat intelligence sharing does not solely depend on the volume of data being shared. Studies show that organizations sometimes "travel on," gaining from shared intelligence without contributing beneficial insights themselves. This imbalance can affect the quality of data being shared, and make it less actionable and potentially overwhelming recipients with irrelevant data.

- **Privacy and Confidentiality Concerns**

There are numerous risks associated with maintaining privacy and confidentiality when sharing intelligence. Considerations regarding the disclosure of sensitive data or breaching privacy laws make organizations reluctant to exchange CTI. These concerns impede the efficacy of ISACs and other CTI forums.

- **Fragmentation and Lack of Interoperability**

Although interoperability is the ultimate objective of standards such as STIX and TAXII, the reality is that many organizations utilize several proprietary formats and forums. This fragmentation may make it more challenging to convey and incorporate data across multiple systems, hindering the seamless interchange of data.

FEDERATED LEARNING (FL) IN THREAT INTELLIGENCE SHARING

Given the challenges associated with conventional threat intelligence sharing standards, modern technological solutions such as Federated Learning (FL) offer a more efficient method for addressing the privacy concerns that hinder the efficient sharing of threat intelligence. FL is a decentralized machine learning approach that necessitates numerous parties to collaborate and build models without disclosing the raw data. To safeguard the confidentiality of the organization's dataset, only model updates are exchanged. With FL, organizations can jointly refine threat detection models while retaining ownership and control of their confidential data, complying with regulations concerning privacy.

In addition, FL is especially advantageous in industries where data is governed by privacy regulations, particularly the fintech and healthcare sectors, with laws such as PCI DSS, HIPAA, and GDPR, which are strict regarding the sharing of sensitive data. Organizations can benefit from collaborative machine learning while conforming to standards.

How Federated Learning Protects Data Privacy

As a distributed machine learning approach, the model training procedure occurs at autonomous and decentralized servers that maintain local data. Each participant (which could be a device belonging to a participating organization) trains the model on their local server and publishes model updates to a central server, thereby eliminating the transmission of private information to a central server. The central server gathers the updates from multiple servers belonging to the numerous participants and builds the global model.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

FL is a Machine Learning approach that Google introduced in 2016. The primary differentiating factor between FL and traditional ML approaches lies in their distributed nature. Likewise, the aggregated model updates can be evaluated using advanced techniques, such as differential privacy, to further strengthen privacy by adding noise to the data, making it harder to pinpoint specific data points.

METHODOLOGY

This section outlines the architecture, dataset presentation, and evaluation methodology for applying an FL-based structure in collaborative threat intelligence sharing. The methodology facilitates the learning of global threat trends while preserving the privacy of sensitive data.

- **FL System Architecture**

This proposed system design empowers multiple stakeholders to jointly train a threat intelligence model, enabling the secure and efficient sharing of data without exposing raw information.

The architecture of an FL comprises the following elements:

- **Participants (Clients/Organizations)**

The participating stakeholder organizations, or their local servers or devices, serve as the client. Each participant harvests their intelligence. Data, featuring IOCs, vulnerability data, TTPs, threat event logs, dark web monitoring data, incident reports, Zero-day threat information, and insider threat indicators, is processed locally. The data is stored within the secure perimeters of the organization. Additionally, differential privacy techniques can be introduced to add randomness to the raw data before it is sent to the central aggregator.

- **Central Aggregating Server**

A secure central server does the training operation. The server initializes the larger workflow, gathers processed data in the form of updates (not the raw data itself) from distributed local servers, and executes model aggregation, typically through Federated Averaging (FedAvg). This aggregation method calculates a weighted average of model parameters from the participating clients. A technique improved by decentralized approaches as proposed by Sun et al. (2021) in their study on Decentralized Federated Averaging.

- **Communication Protocol**

Through secure aggregation, encryption, or cryptographic masking can be applied to the updates sent from the local servers and transmitted through secure communications protocols such as SSL/TLS.

- **Summarized Workflow**

1. A central server launches a global learning algorithm randomly or based on previously acquired intelligence and broadcasts it to selected client devices.

2. Each participating client device receives a copy of the broadcast and locally trains its models using its private datasets. Model training on local servers is typically done through preset intervals or stages.

3. The local training results are a generation of local model updates. The model updates are encrypted and securely transmitted to the central aggregating server.

4. The central server aggregates the updates to produce comprehensive intelligence that can generate global CTI interference, which is exported for further downstream procedures or deployed as CTI artifacts back to the participating organizations' endpoints.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

Workflow Architecture Diagram

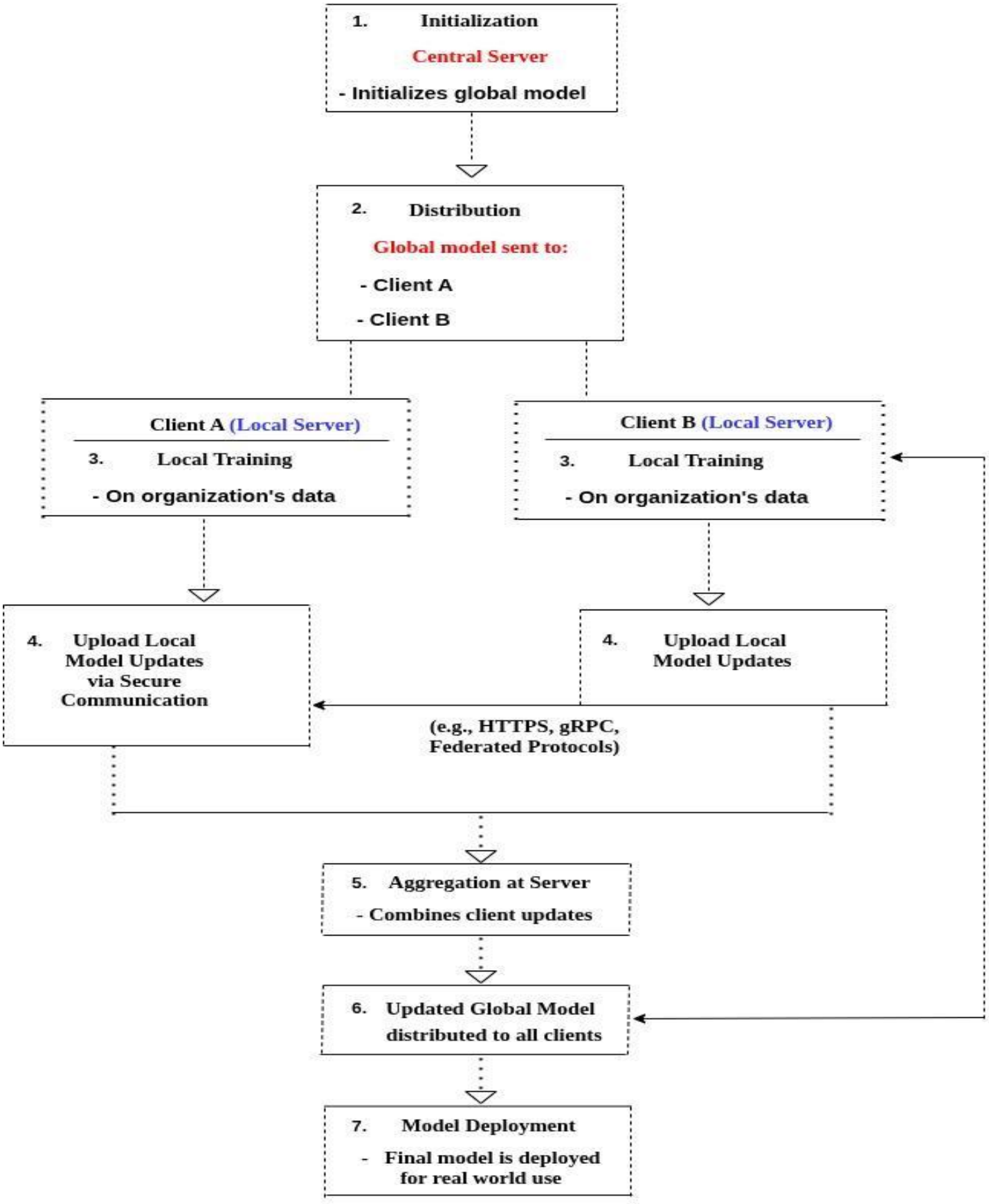

Figure 1: The FL workflow shown above starts with initialization by the remote central model. The central server provides an aggregated global model to participating clients, where model training is performed locally, and model updates are transmitted securely via a communication protocol. Steps 3-6 are executed repeatedly in several communication cycles, upon which the finalized global model is deployed.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

DATASET PREPARATION

The dataset for CTI can be obtained from three primary sources. Data can be obtained through real-life security breaches, simulated environments, or publicly available open-source resources. The categories below encompass various data sources and are organized according to their attributes within the threat ecosystem.

- Indicators of Compromise (IOCs)
- Threat Actor Information.
- Vulnerability and Exploit Intelligence
- Contextual and Strategic Intelligence
- Internal and Organizational Intelligence

- **Data Preprocessing For Federated Learning In Threat Intelligence**

Data processing is a fundamental step in every machine learning pipeline; its value is heightened in federated learning due to the localization of data and the need for confidentiality. Before the teaching begins, data processing ensures that essential, meaningful, and secure logs are obtained from all participating clients.

- **Log Sanitation and Standardization**

Unparsed security logs obtained from devices, firewalls, SIEM, DLP, EDR, and Cloud security solutions often contain incoherent, insufficient, or chaotic entries. In FL structures, the conventional centralized data sanitation is impractical. Alternatively, each client processes data locally using techniques supervised by general criteria, such as data formats and uniform thresholds derived from aggregated statistics.

- **Attribute Retrieval**

Significant information is extracted from security logs to represent notable events. The attribute typically consists of:

1. **Time-Based Data:** Time of the event, day of the week, log ingestion time, etc.
2. **Network Traffic Metrics:** Traffic volumes and dimensions, packet sizes, and network protocols.
3. **Connection Features:** Communication ports, Source/Destination IP addresses.
4. **Behavioural Trends:** Trends of event types, access patterns, etc
5. **Session Metrics:** Session durations, number of failed and successful login attempts

This retrieval is done locally. As described in FedPS, Federated preprocessing tools support this by allowing clients to compute and share summarized statistics relating to protocol frequencies, behavioural trends, and aggregated min/max timestamps that guide homogenous transmission throughout the system (*Anonymous, n.d.*).

- **Event Labeling**

Prior information or threat detection techniques can be used to apply labels to security events. The labels might comprise the following:

- **Malicious or Benign**

Attack Classification - insider threat, ransomware, phishing, malware, brute force, denial of service, web attacks, etc

To ensure label consistency across organizations participating in federated sharing, third-party CTI enrichment or a generally agreed-upon classification may be required.

RESULTS AND DISCUSSION

Detection Accuracy vs. Privacy Trade-offs

The findings in this work illustrate that while global FL models, such as FedAvg, achieve an acceptable level of accurate detection across participating clients, their efficacy is significantly limited in scenarios involving a substantial amount of non-IID data, a prominent attribute in sharing CTI across multiple organizations. Thus,

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue IV April 2026

adopting only global leads to a diminished detection precision for organizations with rare threat trends or uncommon network setups.

The adoption of personalized FL approaches, such as Per-FedAvg and pFedMe, significantly boosts per-client precision by customizing model behavior to local data densities. Nevertheless, these approaches introduce challenges of computational costs, communication issues, and potential data leakage through model updates to central servers, particularly when differential privacy and secure aggregation measures are not implemented. Our evaluation using the Holistic Evaluation Metrics (HEM) framework reveals that personalization can yield a 10% gain in precision, albeit at the expense of extended integration time and reduced equity across participating clients (Sun, Li, & Wang, 2021).

An ideal FL system for CTI should be use-case reliant, combining anonymity and personalization details with organizational attack surface and compliance needs. Furthermore, incorporating differential privacy offers a measurable decrease in model performance, with noise-introduced updates contributing to a 2-6% reduction in accuracy (Kim, Park, & Lee, 2022). However, this tradeoff is acceptable in high-confidentiality scenarios, as guaranteed privacy outweighs any additional performance degradation, especially when the detected threats involve controlled data such as PHI or PII.

LIMITATIONS AND FUTURE WORKS

While this study helps foster the adoption of Federated learning in privacy-preserving cyber threat intelligence sharing, several limitations exist that should guide future studies and FL systems.

- **Limited Support for Vertical FL Scenarios**

The current approach assumes that all clients share identical feature spaces, like in horizontal FL. ISPs, supply chain businesses, software vendors, and cloud service providers are examples of participants in CTI ecosystems that may exhibit similar or dissimilar threat characteristics. The current system cannot account for these vertical FL instances.

Future studies should investigate vertical FL approaches and systems that can incorporate heterogeneous attributes without compromising data privacy.

- **Fairness and Participation Incentives**

Clients with minimal data to contribute, unusual attack types, and overall heterogeneous data quality and volumes may result in some participating clients benefiting less than others in global models, which can dissuade participation (Zhao et al., 2018; Nguyen, Nguyen, & Nguyen, 2024). While metrics such as entropy and Jain's Index contribute to evaluating these gaps, it is less viable to reward clients who do poorly or receive insufficient data, nonetheless.

Future studies could help sustain equity in a multi-organization CTI sharing scenario by adopting strategies such as a federated adaptive aggregation algorithm or reputation-based FL systems.

CONCLUSION

This study highlights the prospects of FL as a feasible and privacy-preserving approach to CTI sharing in a collaborative environment. FL handles primary issues related to confidentiality, compliance, and interoperability, allowing participating organizations to train joint models without exposing confidential information. The outcome demonstrates that personalized models significantly enhance performance for organizations with a unique threat and attack landscape, while global models excel at generalizing across clients. Model evaluation, detection accuracies, and privacy trade-offs are considered. This work presents an elementary framework for secure CTI sharing; however, additional research will be necessary to address some operational limitations and its practicality in real-world CTI sharing use cases.

REFERENCES

1. Abu, M. S., Selamat, S. R., Ariffin, A., & Yusof, R. (2018). Cyber threat intelligence: Issues and challenges. International Journal of Electrical and Computer Engineering (IJECE), 10(1), 371–379. https://doi.org/10.11591/ijeecs.v10.i1.pp371-379
2. Alaeifar, P., Pal, S., Jadidi, Z., Hussain, M., & Foo, E. (2024). Current approaches and future directions for cyber threat intelligence sharing: A survey. Journal of Information Security and Applications, 103786. https://doi.org/10.1016/j.jisa.2024.103786
3. Anomali. (n.d.). Partner datasheet: Carbon Black. https://www.anomali.com/resources/datasheets/partner-datasheet-carbon-black
4. Anonymous. (n.d.). FEDPS: Federated data preprocessing via aggregated statistics. OpenReview. https://openreview.net/pdf?id=eeC1bSkUrY
5. APNIC. (2016, June 24). Cyber threat intelligence sharing: Understanding the technology. https://blog.apnic.net/2016/06/24/cyber-threat-intelligence-sharing-understanding-the-technology/
6. Chakraborty, O., & Boudguiga, A. (2024). A decentralized federated learning using reputation (Report No. 2024/506). Cryptology ePrint Archive. https://eprint.iacr.org/2024/506.pdf
7. Dash, B., Sharma, P., & Ali, A. (2022). Federated learning for privacy-preserving: A review of PII data analysis in FinTech. International Journal of Software Engineering & Applications, 13(4), 1–10. https://doi.org/10.5121/ijsea.2022.13401
8. Divi, S., Lin, Y., Farrukh, H., & Celik, Z. B. (2021). New metrics to evaluate the performance and fairness of personalized federated learning. In FL-ICML Workshop. https://beerkay.github.io/papers/Berkay2021ICMLFLPersonalizedFL.pdf
9. Kim, J., Park, J., & Lee, H. (2022). Privacy-preserving federated learning using homomorphic encryption. Applied Sciences, 12(2), 734. https://doi.org/10.3390/app12020734
10. Li, Y., Ibrahim, J., Chen, H., Yuan, D., & Choo, K. K. R. (2024). Holistic evaluation metrics: Use case sensitive evaluation metrics for federated learning. arXiv preprint arXiv:2405.02360. https://arxiv.org/abs/2405.02360
11. Mohammed Khan, R. H. (n.d.). A comprehensive study on federated learning frameworks.
12. Nigeria Data Protection Regulation. (2019). Nigeria data protection regulation (NDPR). National Information Technology Development Agency (NITDA). https://nitda.gov.ng/wp-content/uploads/2020/11/NigeriaDataProtectionRegulation11.pdf
13. Nguyen, T. D., Nguyen, T. M., & Nguyen, T. T. (2024). Heterogeneity challenges of federated learning for future wireless communication networks. Journal of Sensor and Actuator Networks, 14(2), 37. https://doi.org/10.3390/jsan14020037
14. OpenReview. (n.d.). Federated averaging as expectation-maximization. https://openreview.net/pdf?id=eoQBpdMy81m
15. Pei, F., Xie, Y., Shi, M., & Xu, T. (2025). Adaptive aggregation for federated learning using representation ability based on feature alignment. Knowledge-Based Systems, 113560. https://doi.org/10.1016/j.knosys.2025.113560
16. Stojkovski, B., Koenig, V., Lenzini, G., & Rivas, S. (n.d.). What's in a cyber threat intelligence sharing platform?
17. Sukhabogia, S., & Anusha, M. (2021). A theoretical review on the importance of threat intelligence sharing and the challenges intricated.
18. Sun, T., Li, D., & Wang, B. (2021). Decentralized federated averaging. arXiv preprint arXiv:2104.11375. https://arxiv.org/abs/2104.11375
19. The Johns Hopkins University Applied Physics Laboratory & Bio-ISAC. (n.d.). Going viral: Bioeconomy defense. https://bioisac.org/wp-content/uploads/2021/10/Going-Viral-Bioeconomy-Defense.pdf
20. The MITRE Corporation. (n.d.). Standardizing cyber threat intelligence information with the structured threat information eXpression (STIX™). https://www.mitre.org/sites/default/files/publications/stix.pdf
21. Trocoso-Pastoriza, J. R., Mermoud, A., Bouyé, R., Marino, F., Bossuat, J. P., Lenders, V., & Hubaux, J. P. (2022). Orchestrating collaborative cybersecurity: A secure framework for distributed privacy-preserving threat intelligence sharing. In European Symposium on Research in Computer Security. https://arxiv.org/pdf/2209.02753

22. Wei, K., Li, J., Ma, C., Ding, M., Wei, S., Wu, F., Chen, G., & Ranbaduge, T. (2022). Vertical federated learning: Challenges, methodologies, and experiments. arXiv preprint arXiv:2202.04309. https://arxiv.org/abs/2202.04309
23. Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated learning with non-IID data. arXiv preprint arXiv:1806.00582. https://arxiv.org/abs/1806.00582
24. Zscaler. (n.d.). The ISAC advantage for collective threat intelligence. https://www.zscaler.com/cxorevolutionaries/insights/isac-advantage-collective-threat-intelligence