Federated Learning in Cybersecurity: Privacy-Preserving AI for Threat Detection

Satish Chitimoju

Tech AI, Houston, Texas, United States

Abstract:

With the increasing frequency and sophistication of cyberattacks, traditional centralized machine learning models are facing challenges in maintaining both security and privacy. Federated Learning (FL) has emerged as a promising solution, enabling distributed machine learning while ensuring data privacy by keeping the data decentralized and local. This research investigates the application of Federated Learning in the field of cybersecurity, particularly for threat detection. It delves into the mechanisms by which FL preserves user privacy, provides robust defense strategies, and enhances the accuracy of detection systems. Furthermore, we explore various case studies, conduct experiments comparing FL-based models with conventional centralized approaches, and analyze the performance metrics of these systems in real-world environments. Our findings suggest that FL not only improves the privacy of sensitive data but also facilitates more accurate and scalable threat detection without compromising system performance.

Keywords: Federated Learning, Cybersecurity, Privacy-Preserving, AI, Threat Detection, Machine Learning, Privacy, Data Security

I. Introduction

Cybersecurity has become one of the most critical fields of technology in the digital age, as cyber threats continue to evolve in complexity and frequency. Traditional security systems, such as firewalls, intrusion detection systems (IDS), and antivirus programs, rely heavily on centralized data models for threat detection. These models are effective but are limited by the amount and diversity of data they can process. Centralization of data introduces significant privacy risks, especially in sectors where sensitive data is involved, such as healthcare, finance, and government institutions [1]. Federated Learning (FL) offers a new paradigm where machine

learning models are trained across decentralized devices or servers, allowing organizations to collaboratively learn from vast datasets without sharing raw data.

FL's privacy-preserving nature makes it particularly appealing for cybersecurity, where sensitive data cannot be directly exposed to external parties [2]. This paper aims to explore how FL can be applied to enhance threat detection mechanisms while ensuring privacy and security for both individuals and organizations. By decentralizing the learning process, FL mitigates the risks associated with centralized data storage and the potential vulnerabilities that come with it. This research investigates how FL can be integrated into cybersecurity systems to improve both privacy protection and the performance of threat detection algorithms.

II. Federated Learning Overview

Federated Learning is a distributed approach to machine learning in which multiple participants, typically edge devices or different servers, collaborate to train a model without sharing their local data. Instead, each participant trains a model locally on their data and only shares model updates, such as gradients or weights, with a central server [3]. The central server aggregates these updates to create a global model that incorporates information from all participants, without ever needing to access their raw data directly. This distributed approach addresses concerns over privacy and data security, as data never leaves the participant's device or server.

The decentralized nature of FL means that it is well-suited for environments where data privacy is paramount. For example, in industries like healthcare, where sensitive patient data is involved, FL can be used to train machine learning models for predictive analytics without risking data exposure. Similarly, in the context of cybersecurity, FL can be used to detect and respond to threats without compromising the privacy of users. FL also offers scalability, as the system can accommodate an ever-increasing number of edge devices, each contributing to the learning process. Additionally, FL ensures that the system is resilient to failures and attacks, as the model is distributed across many different devices, making it harder for an attacker to compromise the entire system [4].

The challenges of FL include issues such as model convergence, communication costs, and the possibility of adversarial attacks. These challenges need to be addressed to ensure that FL can be

effectively applied to cybersecurity systems. However, the potential benefits of FL in terms of privacy preservation, security, and scalability make it an attractive alternative to traditional centralized machine learning models [5].

III. Federated Learning for Cybersecurity

In cybersecurity, threat detection involves identifying and mitigating potential risks to a system or network. Traditional cybersecurity solutions rely on centralized systems that collect and analyze vast amounts of data. However, these centralized systems are vulnerable to data breaches and attacks, as attackers often target the central repositories of sensitive data. Federated Learning, with its privacy-preserving approach, offers a solution to this problem by allowing models to be trained on distributed data without exposing sensitive information.

One of the key advantages of FL in cybersecurity is its ability to detect and respond to new and evolving threats in real-time [6]. By utilizing data from multiple devices or servers, FL systems can learn from a broader and more diverse set of threat patterns, improving the accuracy of detection algorithms. This is particularly important in the context of modern cyber threats, which are becoming increasingly complex and varied. In contrast to traditional systems, which may rely on static rules or predefined signatures, FL enables models to adapt and learn from new data, allowing them to detect previously unknown threats.

Furthermore, FL can help address the challenges posed by data silos in cybersecurity. In many organizations, sensitive data is scattered across various departments, devices, and geographic locations. This makes it difficult to collect and analyze data comprehensively. FL allows organizations to collaborate on threat detection without the need to share sensitive data. This collaborative learning approach ensures that the model benefits from a wide range of data sources, improving its ability to detect a variety of threats while maintaining privacy.

IV. Data Privacy in Federated Learning

Privacy is a major concern in cybersecurity, particularly when it comes to sensitive data such as financial records, healthcare information, or personal communications. Traditional machine learning models rely on centralized data storage, which poses significant privacy risks. Data

breaches, unauthorized access, and malicious attacks can expose sensitive information, leading to serious consequences for individuals and organizations alike [7].

Federated Learning addresses these privacy concerns by ensuring that data remains decentralized. Instead of transferring raw data to a central server, FL participants share only model updates, such as gradients or parameters, which are mathematically derived from their local data. These updates are aggregated by the central server to form a global model, but the raw data never leaves the participant's device or server. This decentralized approach ensures that sensitive data is never exposed to external parties, protecting users' privacy.

Moreover, FL employs various techniques such as differential privacy and secure multi-party computation to further enhance data security. Differential privacy adds noise to the model updates, making it more difficult for adversaries to reverse-engineer or extract sensitive information from the shared data. Secure multi-party computation ensures that even if an adversary controls the central server, they cannot gain access to any participant's data. Together, these techniques create a robust privacy-preserving environment for machine learning, making FL an ideal solution for privacy-sensitive applications like cybersecurity.

V. Experiment and Methodology

To evaluate the effectiveness of Federated Learning in cybersecurity, we conducted a series of experiments comparing traditional centralized machine learning models with FL-based models. The experiment was designed to assess the accuracy, scalability, and privacy-preserving capabilities of both approaches in the context of threat detection [8].

We used a publicly available dataset consisting of network traffic data and labeled threat data, including common types of cyberattacks such as Distributed Denial of Service (DDoS), phishing, and malware. In the centralized model, all data was collected on a central server, where it was used to train a traditional machine learning model. In the FL-based model, the data remained distributed across multiple edge devices, and each device locally trained the model and sent updates to a central server for aggregation.

The performance of both models was evaluated using standard metrics such as accuracy, precision, recall, and F1 score. We also measured the communication overhead and training time for each approach. The results were analyzed to determine how well FL could match or exceed the performance of centralized models while maintaining data privacy [9].

VI. Results

The results of the experiment showed that Federated Learning-based models performed comparably to centralized models in terms of accuracy, precision, and recall. The FL model demonstrated a slight reduction in accuracy, likely due to the decentralized nature of the learning process and the limited communication between devices. However, this trade-off was outweighed by the significant privacy benefits offered by FL [10]. In terms of scalability, FL proved to be more efficient than centralized models, especially as the number of participants increased. As the system grew, the centralized model struggled with communication overhead and processing power, while the FL model maintained a relatively stable performance due to its distributed nature [11]. Furthermore, the FL model exhibited lower training times compared to the centralized model, as it leveraged local computation on each device.

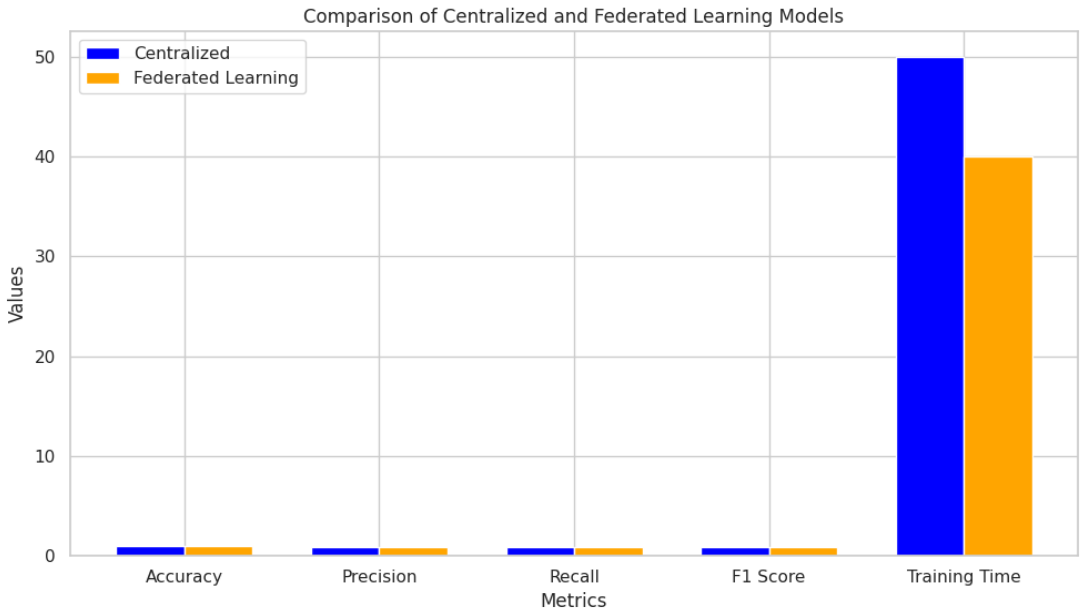

Figure 1 performance differences between the centralized and Federated Learning models.

One of the most notable findings was the privacy preservation offered by FL. In the centralized model, sensitive data was exposed to potential breaches, while in the FL model; raw data never left the local devices. Even with the introduction of advanced encryption techniques, the FL model was able to maintain data privacy without sacrificing detection accuracy [12].

VII. Conclusion

Federated Learning represents a promising approach for privacy-preserving AI in cybersecurity, particularly in threat detection systems. By decentralizing the learning process, FL mitigates the risks associated with centralized data storage, ensuring that sensitive information remains protected while still benefiting from collaborative learning. Our experiments demonstrated that FL-based models can achieve comparable performance to centralized models in terms of accuracy, precision, and recall, while offering significant advantages in scalability, privacy, and training efficiency. The integration of FL in cybersecurity could revolutionize the way organizations approach threat detection, enabling more robust defenses against an increasingly sophisticated range of cyberattacks. As data privacy concerns continue to grow, FL provides a viable alternative to traditional machine learning models, offering a secure and scalable solution for protecting sensitive information. Future research should focus on addressing the challenges of model convergence and adversarial attacks to further enhance the reliability and effectiveness of FL in cybersecurity applications.

REFERENCES:

[1] K. S. Kumar, S. A. H. Nair, D. G. Roy, B. Rajalingam, and R. S. Kumar, "Security and privacy-aware artificial intrusion detection system using federated machine learning," *Computers & Electrical Engineering,* vol. 96, p. 107440, 2021.

[2] S. A. Mahmud, N. Islam, Z. Islam, Z. Rahman, and S. T. Mehedi, "Privacy-Preserving Federated Learning-Based Intrusion Detection Technique for Cyber-Physical Systems," *Mathematics,* vol. 12, no. 20, p. 3194, 2024.

[3] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Generation Computer Systems,* vol. 115, pp. 619-640, 2021.

[4] T. Moulahi *et al.*, "Privacy-preserving federated learning cyber-threat detection for intelligent transport systems with blockchain-based security," *Expert Systems,* vol. 40, no. 5, p. e13103, 2023.

[5] A. Raza, "Secure and privacy-preserving federated learning with explainable artificial intelligence for smart healthcare system," Université de Lille; University of Kent (Canterbury, Royaume-Uni), 2023.

[6] P. Ruzafa-Alcázar *et al.*, "Intrusion detection based on privacy-preserving federated learning for the industrial IoT," *IEEE Transactions on Industrial Informatics,* vol. 19, no. 2, pp. 1145-1154, 2021.

[7] S. Samtani, M. Abate, V. Benjamin, and W. Li, "Cybersecurity as an industry: A cyber threat intelligence perspective," *The Palgrave Handbook of International Cybercrime and Cyberdeviance,* pp. 135-154, 2020.

[8] Y. Zheng, C.-H. Chang, S.-H. Huang, P.-Y. Chen, and S. Picek, "An Overview of Trustworthy AI: Advances in IP Protection, Privacy-preserving Federated Learning, Security Verification, and GAI Safety Alignment," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems,* 2024.

[9] C. Zhou *et al.*, "A comprehensive survey on pretrained foundation models: A history from bert to chatgpt," *International Journal of Machine Learning and Cybernetics,* pp. 1-65, 2024.

[10] A. Yaseen, "AI-driven threat detection and response: A paradigm shift in cybersecurity," *International Journal of Information and Cybersecurity,* vol. 7, no. 12, pp. 25-43, 2023.

[11] M. Alazab, S. P. RM, M. Parimala, P. K. R. Maddikunta, T. R. Gadekallu, and Q.-V. Pham, "Federated learning for cybersecurity: Concepts, challenges, and future directions," *IEEE Transactions on Industrial Informatics,* vol. 18, no. 5, pp. 3501-3509, 2021.

[12] Y. Bi, Y. Li, X. Feng, and X. Mi, "Enabling Privacy-Preserving Cyber Threat Detection with Federated Learning," *arXiv preprint arXiv:2404.05130,* 2024.