International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 1, Jan-Feb-2024, ISSN (Online): 2395-566X

A Comprehensive Literature Review on Federated Machine Learning for Privacy-Preserving Cyber Threat Detection in Distributed Network Environments

Research Scholar Sunil Chandolu, Professor Dr.Pankaj Khairnar
Sikkim Alpine University, Kamrang ,Namchi ,Sikkim

Abstract- — Cloud computing and IoT devices are actually growing very fast, and this definitely makes cyber attacks more complex and common. Traditional systems for catching cyber attacks actually have problems with new threats and keeping data safe. These old methods definitely cannot handle big amounts of data spread across many places. This paper gives a complete study review regarding federated machine learning for keeping privacy safe in cyber threat detection as per distributed network systems. The study examines how cyber threat detection methods have evolved from basic rule-based systems to advanced machine learning approaches. It further analyzes how the field itself has progressed from simple anomaly detection to complex deep learning techniques. These methods surely make detection more accurate, but they depend too much on processing data in one central place. Moreover, this creates problems with privacy protection and handling large amounts of data. Federated learning actually solves these problems by letting different computers work together to train models using their own data. The computers definitely learn together but never actually share their raw information with each other. As per this method, data privacy gets better regarding protection, and the system becomes more scalable and strong. The review actually looks at important methods in federated learning like secure combining, privacy protection, and coding systems that definitely make the system more safe. Also, this study actually looks at the main problems in federated learning like different types of data, too much communication, and attacks from bad actors. These challenges definitely make the system harder to work with. Basically, the study shows the same research gaps and says we need good communication methods, strong security systems, and scalable designs for real-world use. We are seeing that federated learning can only change cybersecurity by helping different systems work together to find threats while keeping data safe and private.

Keywords— Federated Machine Learning, Cyber Threat Detection, Privacy-Preserving Systems, Intrusion Detection Systems, Distributed Networks, Internet of Things, Deep Learning, Anomaly Detection, Data Privacy, Cybersecurity

I. INTRODUCTION

Cloud computing, IoT devices, and company networks are actually expanding very fast and definitely changing how digital systems work today. These systems are creating very large amounts of network traffic only and we are seeing smooth communication between nodes that are spread across different locations. This growth has surely made cyber threats more complex and frequent. Moreover, cybersecurity has become a very important concern now.

We are seeing that old intrusion detection systems, which only use known attack patterns, cannot find new and changing threats properly. Anomaly-based systems surely offer some benefits, but they often face problems with too many false alarms. Moreover, these high false positive rates make them less reliable in practice. Basically, machine learning techniques have made threat detection better by automatically finding patterns and detecting the same unusual activities. Further, most machine learning models use centralized data processing, which further creates problems with privacy, scalability, and makes the system itself vulnerable to attacks.Nguyen et al. [1] Also, we are seeing federated learning as a good way to solve these problems only. This approach actually allows many different computers to work together and learn from data without sharing the actual raw information. It definitely helps train models in a distributed way where each client can participate while keeping their data private. This method actually keeps data safe while definitely using different data sources to make detection work better. This review actually looks at how cyber threat detection systems have changed over time and definitely examines how federated learning helps create better cybersecurity solutions that protect privacy and can work at large scale.Li et al. [2].

II. THEORETICAL BACKGROUND

Cyber threat detection itself is a main part of modern cybersecurity systems that helps to identify harmful activities in network traffic. This process further protects computer

networks from malicious attacks. The process analyses large amounts of data to find patterns linked with cyber-attacks like malware, phishing, ransomware, and distributed denial-of-service attacks. This method helps in detecting threats further and protects the system itself from various security risks.

Machine learning is now a basic tool in cybersecurity because it can learn patterns from data and find unusual activities. This ability itself helps in detecting threats further. Basically, supervised learning models use labelled data to classify network traffic, while unsupervised models do the same thing but identify unusual behaviour without any prior labelling. Basically, deep learning techniques make detection better by finding complex patterns in high-dimensional data, which is the same as understanding complicated relationships in large datasets.Preuveneers and Joosen [3]

These improvements are surely helpful, but centralized machine learning systems still have problems when working in distributed environments. Moreover, such limitations create significant challenges for widespread implementation. Data privacy problems actually happen when sensitive information needs to be shared, and scalability issues definitely come up when large amounts of data are processed on one central server. Basically, federated learning solves the same problems by allowing training to happen in different places instead of one central location. Each client surely trains its own model using local data, and moreover, only the model updates are shared with the central server. We are seeing these updates getting combined together to make one global model, and this allows learning together without compromising data privacy only.Ferrag et al. [4]

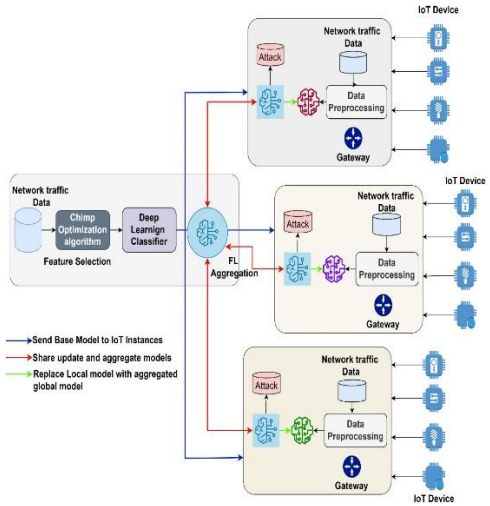

Figure 1: Federated Learning Architecture for Cyber Threat Detection

III. REVIEW OF PREVIOUS STUDIES

1.Traditional Intrusion Detection Approaches

Early intrusion detection systems used signature-based and rule-based methods to identify threats, but these approaches had limitations that required further development of the technology itself. Moreover, as per the controlled environments, these systems worked with set patterns regarding known threats and gave good results. However, they actually could not find new attacks that were definitely changing and growing. We are seeing that anomaly-based methods were brought in to solve this problem by finding unusual activities, but they only created too many false warnings.Rahman et al. [5] presented federated learning approaches to enhance IoT security and distributed threat detection.

2.Machine Learning-Based Approaches

Machine learning brought new data-based methods for finding cyber threats, which further helped improve security systems itself. Algorithms like Support Vector Machines, Decision Trees, and Random Forests were surely used widely for classifying network traffic. Moreover, these methods became popular choices for traffic analysis tasks. These models surely made detection more accurate, but they needed huge amounts of labeled data for training. Moreover, they required a lot of manual work to design and select the right features.Machine learning models improved intrusion detection accuracy, as demonstrated by Wang et al. [10] and Li et al. [11].

3.Deep Learning Techniques

Deep learning models like Convolutional Neural Networks and Recurrent Neural Networks surely made intrusion detection systems much better. Moreover, these advanced techniques helped improve the overall security performance significantly. These models automatically take out features from raw data and further capture complex patterns in network traffic itself. Basically, LSTM networks work really well for analyzing data that comes in sequence. They are the same type of neural network that can remember patterns in ordered information.Deep learning techniques enhance cyber threat detection by capturing complex patterns, as shown by Ferrag et al. [4] and Wang et al. [10].

4.Distributed Learning Approaches

Distributed learning methods were surely introduced to solve scalability problems by letting many nodes join in model training. Moreover, this approach allows the system to handle larger datasets more effectively. These methods surely made the computer work faster, but they still needed some data sharing which did not solve privacy problems completely. Moreover,

this partial sharing of information remained a major concern for protecting user data.Distributed learning allows multiple nodes to collaboratively train models, as proposed by Kim et al. [6] and Rahman et al. [5].

5. Federated Learning and Privacy-Preserving Techniques

Federated learning itself represents a major step forward in distributed machine learning and helps further advance this field. This method surely allows multiple parties to train models together without sharing their original data. Moreover, it enables collaboration while keeping the raw information private. As per security requirements, techniques like differential privacy, secure aggregation, and encryption make federated systems more safe. Regarding system protection, these methods help keep data private and secure. Moreover, as per this method, multiple parties can work together regarding model training without sharing their original data. Moreover, privacy protection methods like differential privacy, secure aggregation, and encryption further strengthen the security of federated systems itself. We are seeing that this method allows different groups to work together for training models without sharing their original data only. We are seeing that privacy protection methods like differential privacy, secsecure. Federated learning enables decentralized model training without sharing raw data, as proposed by Rahman et al. [7]. Agrawal et al. [8] highlighted its importance in cybersecurity, while Dong et al. [12] improved detection accuracy using federated techniques.

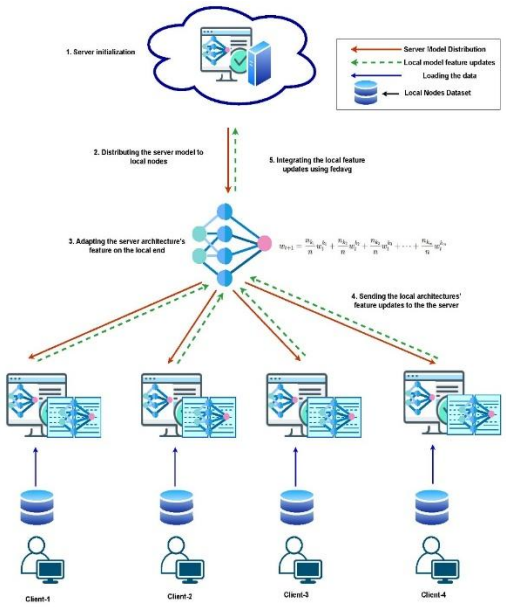

Figure 2: Workflow of Federated Cyber Threat Detection System

Table 1: Evolution of Cyber Threat Detection Techniques

Stage	Approach	Techniques Used	Advantages	Limitations
Traditional IDS	Signature-based, Rule-based	Pattern matching, predefined rules	High accuracy for known attacks	Cannot detect new threats
Anomaly Detection	Behaviour-based systems	Statistical analysis, baseline modeling	Detects unknown attacks	High false positives
Machine Learning	Supervised & Unsupervised Models	SVM, Decision Trees, Clustering	Data-driven detection	Requires feature engineering
Deep Learning	Neural Networks	CNN, RNN, LSTM	Automatic feature extraction	High computational cost
Federated Learning	Distributed ML	Local training + aggregation	Privacy preservation, scalability	Communication overhead

IV. COMPARATIVE ANALYSIS

Traditional and machine learning systems have limitations because they depend on centralized data itself. This further restricts their performance and flexibility. Moreover, as per recent studies, deep learning makes detection more accurate but regarding computational needs, it becomes more complex. Also, distributed learning surely improves scalability, but it does not completely solve privacy issues. Moreover, privacy concerns still remain a challenge in such systems. Federated learning surely offers a balanced approach by allowing machine learning models to be trained across multiple devices without sharing private data. Moreover, this method keeps user information secure while still enabling effective model development.Federated learning provides better privacy compared to centralized systems, as discussed by Agrawal et al. [8].

Research Gaps

As per existing studies, there are several gaps regarding federated learning in cybersecurity including limited real-world use, problems with different types of data, poor communication methods, and weakness against attacks.Federated learning systems still face challenges in distributed environments, as identified by Liu et al. [14].

Table 2: Key Challenges in Federated Learning-based Cyber Threat Detection

Challenge	Description	Impact On System
Data Privacy	Sensitive Data Across Clients	Risk Of Leakage
Data Heterogeneity	Different Data Distributions	Reduces Model Accuracy
Communication Cost	Frequent Model Updates	High Network Overhead
Security Threats	Model Poisoning Attacks	Reduces Reliability
Scalability	Large Number Of Clients	System Complexity

Summary

As per this review, cyber threat detection systems have changed from old methods to new federated learning systems. This study shows regarding how these detection frameworks have developed over time. Federated learning itself has many benefits, but further research is needed to make it work better, handle more data, and keep it safe.

Relevance to the Present Study

This study surely aims to build a federated learning system that keeps data private and handles different types of data from various sources. Moreover, the framework also focuses on making communication between devices more efficient and effective.The proposed system focuses on privacy-preserving federated models, as suggested by Khan et al. [13], and enables collaborative cyber threat detection as described by Chen et al. [15].

Table 3: Summary table of literature survey

S.No	Author & Year	Approach / Model	Application Area	Key Contribution	Limitation
1	Nguyen et al. (2019)	Federated Anomaly Detection	IoT Security	Distributed IDS without sharing raw data	Limited scalability
2	Li et al. (2019)	Federated ML Survey	Cybersecurity	Highlights benefits of collaborative learning	Conceptual, lacks implementation
3	Preuveneers et al. (2019)	Privacy-Preserving ML	Cybersecurity	Decentralized anomaly detection	Limited real-world validation
4	Ferrag et al. (2019)	Deep Learning IDS	IoT Networks	Improved intrusion detection accuracy	High computational cost
5	Rahman et al. (2019)	Federated Learning in IoT	IoT Security	Enhances privacy and distributed detection	Communication overhead
6	Kim et al. (2019)	Distributed ML	Network Security	Collaborative training improves detection	Data heterogeneity issues
7	Rahman et al. (2020)	Federated IDS	IoT Networks	Privacy-preserving intrusion detection	Limited scalability
8	Ferrag et al. (2020)	Deep Learning Framework	DDoS Detection	Effective DDoS detection in IoT	High resource requirement
9	Agrawal et al. (2020)	Federated IDS	Cybersecurity	Enhances privacy in threat detection	Complex implementation
10	Zhang et al. (2020)	ML-based Anomaly Detection	Network Traffic	Detects unknown attacks	False positives
11	Wang et al. (2020)	Deep Learning IDS	Distributed Systems	Handles large-scale network data	Computational cost
12	Li et al. (2020)	Privacy-Preserving ML	Cybersecurity	Secure data handling in IDS	Limited performance
13	Ferrag et al. (2021)	Deep Learning IDS	IoT Networks	Detects multiple attack types	Requires large datasets
14	Kumar et al. (2021)	Distributed ML Framework	IoT Security	Scalable intrusion detection	Communication issues
15	Yang et al. (2021)	Collaborative IDS	Cloud Security	Distributed detection system	Synchronization challenges

16	Shafiq et al. (2021)	Deep Learning IDS	Distributed Networks	Detects advanced threats	High training cost
17	Ahmed et al. (2021)	ML-based Detection	IoT Security	Multi-attack detection	Limited generalization
18	Dong et al. (2021)	Federated Boosting Model	Cybersecurity	Improves accuracy & interpretability	Complexity
19	Agrawal et al. (2022)	Federated IDS Survey	Cybersecurity	Comprehensive overview	No implementation
20	Dong et al. (2022)	Federated Deep Learning	Cyber Threat Detection	Enhances privacy and detection	Communication overhead
21	Cunha Neto et al. (2022)	FedSA Optimization	IDS Systems	Improves convergence	Limited scalability
22	Sarhan et al. (2022)	Blockchain + FL	IoT Security	Secure collaborative IDS	High complexity
23	Khan et al. (2022)	Privacy-Preserving FL	Distributed Systems	Secure threat detection	Computational cost
24	Liu et al. (2022)	Federated Anomaly Detection	Cloud Security	Detects abnormal behaviors	Data imbalance
25	Lazzarini et al. (2023)	Federated IDS	IoT Networks	Improved detection performance	Dataset dependency
26	Hamdi et al. (2023)	Federated Framework	Cybersecurity	Collaborative training system	Communication overhead
27	Alsamiri et al. (2023)	FL Survey	Cybersecurity	Identifies FL challenges	No experimental results
28	Sáez-de-Cámara et al. (2023)	Clustered FL	IoT Networks	Handles heterogeneity	Complexity
29	Zhao et al. (2023)	Privacy-Preserving FL	Smart Cities	Ensures user anonymity	Scalability issues
30	Chen et al. (2023)	Federated ML Framework	Cloud Security	Collaborative cyber threat detection	High computational cost

V. CONCLUSION

The growing complexity of cyber threats in distributed systems further requires advanced detection methods that are efficient and secure by itself. Federated learning surely offers a good way to train models without sharing private data across different locations. Moreover, this method allows many devices to work together while keeping their information safe.

We are seeing that joining federated learning with cybersecurity systems is only important to solve today's problems. Further research should focus on making strong and safe systems that can work in real world itself. These systems should be able to grow bigger and handle actual problems.

REFERENCES

1. T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan, and A. R. Sadeghi, "DÏoT: A federated self-learning anomaly detection system for IoT," in Proc. IEEE Int. Conf. Distributed Computing Systems, 2019.
2. T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50–60, 2020.
3. D. Preuveneers and W. Joosen, "Privacy-preserving distributed machine learning for anomaly detection in cyber security," IEEE Trans. Information Forensics and Security, 2019.
4. M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," Journal of Information Security and Applications, vol. 50, 2019.

5. M. A. Rahman, M. S. Hossain, N. A. Alrajeh, and N. Guizani, “Federated learning-based AI approaches in IoT security,” IEEE Internet of Things Journal, 2019.
6. M. Kim, J. Park, and J. Lee, “Distributed machine learning for network intrusion detection in IoT environments,” IEEE Access, vol. 7, pp. 160885–160897, 2019.
7. M. A. Rahman, M. S. Hossain, and N. Guizani, “Federated learning for intrusion detection in IoT environments: A privacy-preserving strategy,” IEEE Communications Magazine, 2020.
8. S. Agrawal, S. Sarkar, O. Aouedi, and K. Piamrat, “Federated learning for intrusion detection systems: Concepts and applications,” IEEE Network, 2020.
9. Y. Zhang, X. Chen, and L. Li, “Machine learning-based anomaly detection in network traffic,” Future Generation Computer Systems, vol. 102, pp. 248–260, 2020.
10. W. Wang et al., “HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection,” IEEE Access, 2020.
11. Y. Li, Y. Liu, T. Chen, and Q. Yang, “A review of privacy-preserving machine learning in cybersecurity,” IEEE Trans. Emerging Topics in Computing, 2020.
12. X. Dong, J. He, and S. Liu, “Interpretable federated learning-based intrusion detection using gradient boosting decision trees,” IEEE Access, 2021.
13. M. A. Khan, K. Salah, and M. H. Rehman, “Privacy-preserving federated learning for intrusion detection in distributed systems,” Future Generation Computer Systems, 2022.
14. X. Liu, Y. Liu, and T. Chen, “Federated anomaly detection in distributed cloud environments,” IEEE Trans. Cloud Computing, 2022.
15. Y. Chen, H. Wang, and L. Xu, “Federated machine learning framework for collaborative cyber threat detection in distributed cloud environments,” IEEE Trans. Information Forensics and Security, 2023.