GLIMPSE - Journal of Computer Science • Vol. 3(1), JANUARY-JUNE 2024, pp. 4-9

NETWORK INTRUSION DETECTION SYSTEM USING FEDERATED LEARNING

Amrita Bhatnagar
Assistant Professor, Ajay Kumar Garg Engineering College, Ghaziabad, U.P, India
bhatnagaramrita@akgec.ac.in

***Abstract*—Network device security and privacy are greatly enhanced by (IDS). Due to their excellent classification accuracy, Machine Learning (ML) and Deep Learning (DL) with IDS have become increasingly popular. However, the requirement of storing and transmitting data to a centralized server may threaten privacy and security. The primary difficulty facing an anomaly-based intrusion detection system is this. A challenge with anomaly-based IDS is data training. This paper addresses the application of federated learning (FL) in several facets of anomaly detection and offers a federated learning solution to this problem. A discussion of the main implementation issues for FLs is also included, giving insight into the potential areas of future research.**

***Keywords*—Federated, Intrusion, Global, Local, Aggregation**

I. INTRODUCTION

The Google research team introduced federated learning (FL), a distributed system that can simultaneously expand data and protect privacy. FL was presented in 2016. Each workstation's private data is kept private, and it treats businesses, factories, or edge devices with data as clients that take part in federated learning. Customers use their private datasets for training and develop the same deep local models. Make anaggregate model with the same structure as the private model on the cloud center server. They transfer these models using continuous connections between multiple training clients and the central server. Ultimately, a particularly effective aggregate model is jointly constructed to meet specific learning goals.

II. RELATED WORK

The intrusion detection system is a system that can find out malicious data on network traffic and it can also find out intrusion detection on a particular host [1]. It can have two types of detection methods signature-based and anomaly-based IDS. The signature-based intrusion detection system can find out the attack on the basis of some patterns of previous attacks which is stored in the database but it cannot find the Nobel attacks so for noble attacks we can use anomaly-based intrusion detection method which is based on ML /DL techniques [2]. ML/DL techniques in IDS facing major issues of data training. Thedata of the client is not secure because it is sent over the network. So, there is a federate technique for resolving this problem of the Anomaly Intrusion Detection system [3].It can be implemented successfully for network intrusion detection systems for IOT devices[4].

III. INTRUSION DETECTION SYSTEM

An Intrusion Detection System (IDS) plays a critical role in detecting and mitigating cyber-attacks. It is a security technology that monitors network and system activities for signs of unauthorized, malicious, or suspicious behavior. The primary purpose of an IDS is to identify and respond to potential security breaches in real-time or near-real-time, helping to protect computer networks and systems from various types of cyber-attacks [5]. Here's how an IDS contributes to defending against cyber-attacks.

Threat Detection: IDS monitors network traffic and system logs to identify unusual patterns, activities, or anomalies that might indicate a cyberattack. It can detect known attack signatures or deviations from normal behavior, such as increased traffic, unauthorized accessattempts, or unexpected changes in system configurations.

Early Warning: By detecting attacks as they occur or shortly after they start, an IDS provides an early warning system that allows security teams to respond quickly and effectively[6]. This can help minimize the potential damage caused by an ongoing cyber-attack.

Real-time Monitoring: IDS continuously monitors network traffic and system activities, providing real-time insights into potential security incidents. This proactive approach enables timely response and reduces the window of opportunity for attackers.

Incident Response: An intrusion detection system (IDS) can produce alerts or notifications to notify security personnel of potentially dangerous activities. This makes it possible for security teams to look into cyberattacks and act quickly to lessen their effects.

Behavior Analysis: Modern IDS solutions often employ behavioral analysis techniques, which involve learning thenormal patterns of network and system behavior. This enables the system to detect anomalies and deviations from the established baseline, even if the attack doesn't have a known signature.

Network Segmentation: IDS can help enforce network segmentation by monitoring traffic between different segments and detecting any unauthorized communication attempts between them. This helps contain and limit the lateral movement of attackers within a network.

Compliance and Regulation: Many industries and organizations are subject to regulatory requirements that mandate the use of intrusion detection systems for security and compliance purposes. IDS helps demonstrate adherence to these standards.

Forensic Analysis: IDS logs and data can be used for post-incident analysis and forensic investigations. They provide valuable insights into the nature of an attack, its origin, and the compromised systems, helping organizations understand the attack's impact and plan for better defense strategies.

Tuning and Improvement: IDS systems can be fine-tuned over time based on the analysis of detected incidents. This process improves the accuracy of detection and reduces false positives, making the system more effective at identifying genuine threats.

An Intrusion Detection System is a crucial component of a comprehensive cybersecurity strategy, helping organizations identify, respond to, and mitigate cyber-attacks before they cause significant harm[7,8]. It complements other security measures and tools to create a robust defense against evolving cyber threats.

IV. REQUIREMENT OF FEDERATED LEARNING IN IDS

FL is a ML technique that keeps data local and decentralized while enabling numerous participants (servers or devices) to work together to build a common ML mode[9,10]l. When used in conjunction with intrusion detection systems (IDS), this idea has the potential to provide the following benefits:

Privacy Preservation: In the context of IDS, sensitive network and system data often reside within different organizational or network segments

Data Diversity: Different network segments or organizations may have varying network patterns and threat profiles. Federated Learning allows each segment to train a model using its unique data, leading to a more diverse and representative model that can better detect threats across the entire system

Reduced Data Transfer: Traditional centralized approaches in IDS often involve transferring large amounts of data to a central server for analysis.

Efficient Model Updates: In IDS, threat landscapes can change rapidly. Federated Learning enables incremental model updates across segments, allowing each segment to adapt its model to new threats while preserving the existing knowledge of the system.

Decentralized Architecture: IDS with federated learning can operate with a decentralized architecture, making it more resilient to single points of failure and reducing the impact of system disruptions.

Collaborative Knowledge Sharing: Participating segments can collectively contribute to improving the accuracy and effectiveness of the IDS model. Federated Learning facilitates the exchange of insights and model improvements without sharing sensitive raw data.

Regulatory Compliance: In regulated industries, where data sharing is restricted, federated learning can help organizations comply with data protection regulations while still benefiting from a collaborative security approach.

Scalability: Federated Learning can be particularly useful in large-scale environments where a centralized approach might not be feasible due to resource limitations or network constraints.

V. FEDERATED LEARNING ARCHITECTURE

The architecture of federated learning involves several key components and processes that allow this collaborative training to take place. Here's an overview of the typical federated learning architecture:

Central Model: The original model that is sent to client devices for on-site training is known as the central model. It functions as the foundation for group learning. Usually, the central model is an existing model or one that has been pretrained on a larger dataset.

Client Model Training: Each client device trains its local model using its local data. Local training involves iteratively updating the model parameters based on the local data while keeping the model architecture fixed.

Model Update Aggregation: After local training, client devices send their model updates (model parameter changes) back to the server. The server aggregates these updates to create a global model update that reflects the combined knowledge from all participating devices.

Aggregate Model Update: The aggregated model update is applied to the central model, resulting in an updated global model. This global model embodies collective learning from all client devices without directly exposing their raw data.

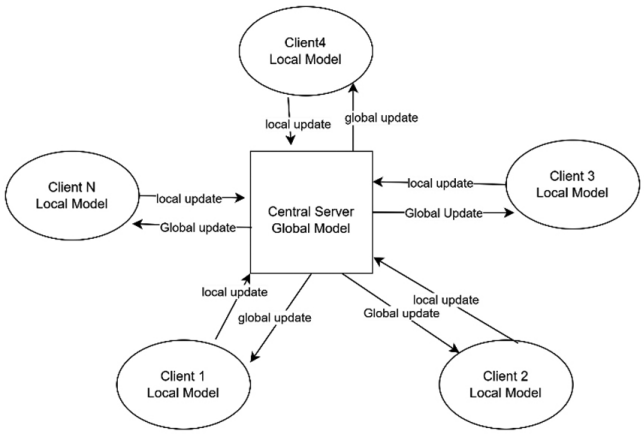
Differential Privacy: In some federated learning implementations, differential privacy techniques are applied to further enhance data privacy. These techniques introduce noise or randomness to the model updates before aggregation, making it harder to infer individual data from the updates.

Communication Protocols: The communication between client devices and the server is a critical aspect of federated learning. Secure and efficient communication protocols are used to exchange model updates, aggregate them, and distribute the updated model.

Model Evaluation and Validation: Periodically, a subset of client devices may evaluate the global model's performance using their local validation datasets. This helps ensure that the global model maintains its effectiveness across different segments.

Model Deployment: Once the global model has been updated and validated, it can be deployed back to the client devices for further local training iterations.

Iterative Process: Federated learning is an iterative process, with multiple rounds of local training, model update aggregation, and global model updates. Over time, the global model becomes more accurate and robust.



VI. FEDERATED LEARNING TYPES

Horizontal and vertical federated learning are two variants of the federated learning paradigm, each addressing different aspects of data distribution and collaboration. Let's explore both concepts:

Horizontal Federated Learning:
In horizontal federated learning, multiple entities or devices possess the same types of data but are collected from different sources. These sources could be different geographical locations, departments within an organization, or various devices.

Vertical Federated Learning: In vertical federated learning, different entities or devices possess different types of data, often related to the same individuals or entities. Each participant holds a subset of features or attributes of a larger dataset, and the goal is to collaboratively learn a model that utilizes the combined information from these sources.

Key features of vertical federated learning: Different Data Types: Participants possess different types of data or attributes related to the same individuals or entities.

Complementary Information: Vertical federated learning combines distinct data sources to create a more comprehensive model, leveraging the strengths of each participant's data. Joint Analysis: The goal is to perform joint analysis without exposing the full data to all participants, preserving data privacy.

Secure Collaboration: Vertical federated learning enables collaboration between parties with different data sources without sharing sensitive information.
An example of vertical federated learning could be combining financial data from different banks to build a credit scoring model. Each bank may have unique attributes about its customers, and by collaboratively training a model, a more accurate credit scoring system can be created.

VII. EXISTING SYSTEM OF IDS

The figure shows the methodology for existing hybrid IDS where signature, as well as anomaly detection techniques, will be used. If the attack occurs then IDS will check the signature in the database if it is available then it will take action for the attack. If the attack pattern is not available in the database it will go for Anomaly detection using machine learning techniques.

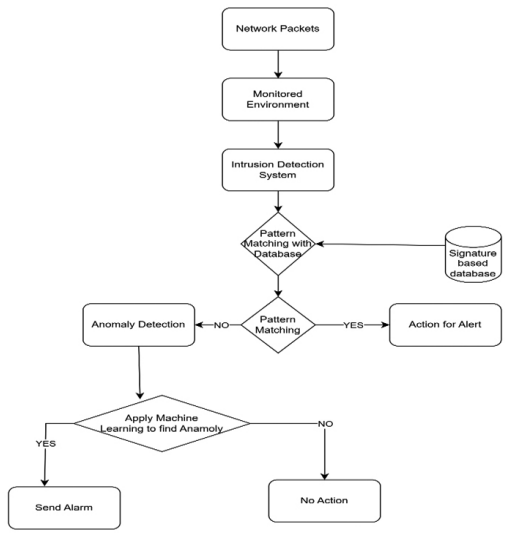
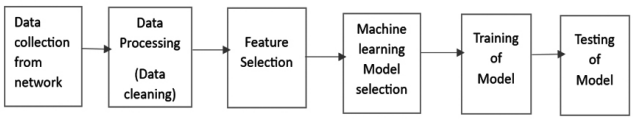


Fig1. Existing System of IDS

VIII. METHODOLOGY FOR ANOMALY DETECTION

For Anomaly detection System we can use these steps to use machine learning.

1. Data Collection
2. Data Preprocessing
3. Feature Selection
4. Training of data
5. Testing of data



IX. CHALLENGES IN DATA TRAINING

Data Quality and Noise: Network data can be noisy and contain anomalies that are not necessarily attacks. Such noise can confuse the IDS and lead to false positives. Additionally, data collected from different sources or with varying quality levels can impact the generalization of the model.

Data Privacy and Ethics: IDS typically rely on collecting and analyzing network traffic data, which can raise privacy concerns. Balancing the need for accurate detection with respecting user privacy is a critical challenge.

X. PROPOSED IDS USING FEDERATED LEARNING

FL can be used in Intrusion detection Systems for training the model. There can be different types of clients on the network and each has its local intrusion detection model.

There will be a centralized global model also. These models can be used with deep learning techniques like CNN, and RNN. There are some steps to complete the data training.

1. Global Model Generation
2. Transfer the Global model parameters to all client side
3. Train and improve the model with client-side data
4. Send updated model parameters to the global model
5. Aggregate the local model parameter
6. After Aggregation send back the model to client-side
7. Now use new parameters
8. Repeat steps 4, 5, 6, 7

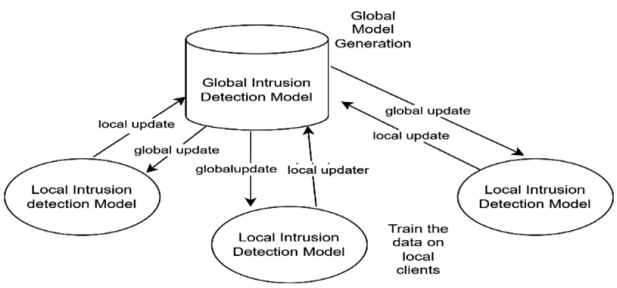


Fig. 2: Federated learning based ids

XI. ADVANTAGES OF FL-BASED IDS

Federated Learning-based Isntrusion Detection Systems (IDS) offer several advantages compared to traditional centralized IDS approaches. These advantages stem from the decentralized and privacy-preserving nature of federated learning. Here are some key benefits:

Collaborative Learning: Federated learning enables different segments of a network or different organizations to collaborate on training a shared IDS model. This collaborative approach allows the IDS to benefit from diverse and distributed knowledge while respecting data privacy constraints.

Enhanced Accuracy and Robustness: Federated learning leverages data from various sources, resulting in a more accurate and robust IDS model. By aggregating insights from different network segments or organizations, the IDS becomes more adept at detecting a wide range of threats and anomalies.

Real-time Response: Federated learning-based IDS can provide real-time or near-real-time detection and response capabilities. Local devices can continuously update their models based on their unique data, leading to quicker adaptation to evolving threat landscapes.

Decentralized Resilience: Federated IDS systems are inherently more resilient to single points of failure or attacks on a central server. The decentralized architecture ensures that disruptions in one segment do not impact the overall system's functionality.

Regulatory Compliance: In industries with strict data protection regulations, federated learning-based IDS can help organizations adhere to compliance requirements while still benefiting from advanced threat detection capabilities.

Scalability: Federated learning can scale effectively across large and distributed networks. New segments can be added to the federation without requiring a complete overhaul of the system's architecture.

Adaptation to Local Context: Localized model training allows each segment to adapt its IDS to the specific threat landscape and patterns it faces. This flexibility improves the system's ability to detect region-specific or segment-specific threats.

Continual Learning: The iterative nature of federated learning allows the IDS to continually improve and adapt over time. As new threats emerge or network conditions change, the model can be updated without disrupting the entire system.

Easier Collaboration: Federated learning fosters collaboration between different entities or organizations without requir-

ing the sharing of sensitive data. This can lead to knowledge exchange and improved collective cybersecurity practices.

Federated learning-based IDS offers a privacy-preserving and collaborative approach to intrusion detection, enabling organizations to benefit from shared threat intelligence while maintaining the security and privacy of their data.

XII. CHALLENGES

While Federated Learning-based Intrusion Detection Systems (IDS) offer many benefits, they also come with their own set of challenges and complexities.

Heterogeneous Data: Different network segments or organizations might have varying data distributions, formats, and quality. Aggregating and reconciling these diverse data sources for effective model training and aggregation can be challenging.

Model Aggregation: Aggregating model updates from various segments while ensuring model integrity and accuracy is a complex task. Balancing the contributions from different segments without negatively impacting the global model's performance is a challenge.

Data Imbalance: Some segments may have significantly more data than others, leading to imbalanced contributions during model aggregation. This can affect the fairness and accuracy of the resulting global model.

Local Model Quality: The quality of local models on client devices can vary due to factors such as limited computational resources, noisy data, or inadequate training. Low-quality local models can negatively impact the overall system's performance.

Model Poisoning Attacks: Adversaries may attempt to inject malicious or incorrect model updates into the aggregation process to degrade the performance of the global model. Detecting and mitigating such attacks is a critical concern.

Central Server Bottleneck: While federated learning aims to distribute training, the central server's role in model aggregation and management can become a performance bottleneck as the number of client devices increases.

Federated Optimization Challenges: Federated learning involves solving optimization problems across distributed devices with limited communication and computation capabilities. Developing efficient algorithms for federated optimization is a research challenge.

Model Drift and Concept Drift: Over time, the threat landscape and data distribution can change, leading to model drift or concept drift. Detecting and adapting to these changes without frequent centralized updates is a challenge.

Lack of Standardization: Federated learning is a relatively new and evolving field, and there's a lack of standardized protocols, frameworks, and best practices. This can lead to interoperability issues and implementation complexities.

Regulatory and Legal Considerations: Managing data privacy and compliance with regulations across different segments or organizations can be complex, especially when sharing model updates that might still contain sensitive information.

Model Fairness and Bias: Federated learning can inherit biases present in local datasets. Ensuring fairness and addressing biases across different segments is a challenge.

Addressing these challenges requires a combination of research, algorithmic development, secure communication protocols, privacy-preserving techniques, and collaborative efforts from the research and industry communities. While federated learning-based IDS has the potential to revolutionize intrusion detection while preserving privacy, careful consideration and solutions are needed to overcome these obstacles.

CONCLUSIONS

In summary, federated learning is a technique that, through data training and local model aggregation, enhances the security and performance of anomaly intrusion detection systems. However, federated learning is not without its challenges, such as the potential for a poisonous attack on local sides that could corrupt global models. Further research is needed to address these kinds of issues.

FUTURE DIRECTIONS

Use of Encryption in Federated Learning

FL benefits greatly from the use of encryption in terms of security and privacy. It ensures that only authorized parties can access and process the data, protecting sensitive data during the communication and aggregation processes.

Edge Computing with FL-based IDS

Edge computing enables local processing of data on devices or edge servers, reducing the need to transfer large amounts of data to a central server. In FL, this means that only model updates or aggregated information need to be transmitted, minimizing communication overhead and latency.

REFERENCES

[1] Thakkar, Ankit, and Ritika Lohiya. "A Review on Challenges and Future Research Directions for Machine Learning-Based Intrusion Detection System." Archives of Computational Methods in Engineering (2023): 1-25.

[2] Khraisat, A., Gondal, I., Vamplew, P. et al. Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecur 2, 20 (2019). https://doi.org/10.1186/s42400-019-0038-7

[3] Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. Knowledge-Based Systems,216,106775. https://doi.org/10.1016/j.knosys.2021.106775

[4] Agrawal, Shaashwat, et al. "Federated learning for intrusion detection system: Concepts, challenges and future directions." Computer Communications (2022).

[5] Rashid, M. M., Khan, S. U., Eusufzai, F., Redwan, M. A., Sabuj, S. R., & Elsharief, M. (2023). A Federated Learning-Based Approach for Improving Intrusion Detection in Industrial Internet of Things Networks. Network, 3(1), 158-179. https://doi.org/10.3390/network3010008

[6] Bag, S. Federated Learning—A Beginners Guide. 15 May2021.Available online: https://www.analyticsvidhya.com/blog/2021/05/federated-learning-a-beginners-guide

[7] Ahmad, R.; Alsmadi, I. Machine Learning Approaches to IoT Security: A Systematic Literature Review. Internet Things 2021, 14, 100365.

[8] Christopher Regan, Mohammad Nasajpour, Reza M. Parizi, Seyedamin Pouriyeh, Ali Dehghantanha, Kim-Kwang Raymond Choo,Federated IoT attack detection using decentralized edge data,Machine Learning with Applications,Volume 8,2022,100263,ISSN 2666-8270

[9] Abdulganiyu, Oluwadamilare Harazeem, Taha Ait Tchakoucht, and Yakub Kayode Saheed. "A systematic literature review for network intrusion detection system (IDS)." International Journal of Information Security (2023): 1-38.

[10] M. H. Bhuyan, D. K. Bhattacharyya and J. K. Kalita, "Network Anomaly Detection: Methods, Systems and Tools," in IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 303-336, First Quarter 2014, doi: 10.1109/SURV.2013.052213.00046.

ABOUT THE AUTHOR

Amrita Bhatnagar works as an assistant professor at the Ghaziabad campus of AKGEC Engineering College. She holds an MTech from BIT Mesra Rachi and a BE from SRMSCET Bareilly. Digital image processing, cyber security, information security, and Python are among the topics she studies. She has numerous articles on information security published in reputable journals.