Federated Learning and Explainable Hybrid Deep Learning for Cyber-Attack Detection in Distributed Big Data Environments: A Comprehensive Survey

Raghunath Kumar Babu D[1, 2], Shashi Kant Gupta [3]

[1] Postdoctoral Researcher, Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia;

[2] Assistant Professor, Department of CSE(CS), Rajeev Gandhi Memorial College of Engineering and Technology, Nandyala, India;

[3] Professor, Department of CSE, Lincoln University College, Petaling Jaya, Selangor, Malaysia

[1, 2] pdf.Raghunath@lincoln.edu.my, [3] shashigupta@lincoln.edu.my

Abstract: The proliferation of distributed computing environments, including cloud computing, IoT networks, and edge systems, has increased the cyber-attack surface and created new challenges in security. Centrally located intrusion detection systems (IDSs) have shortcomings with data privacy, scalability, communication, and transparency. Federated Learning (FL) is a distributed learning system that allows system nodes to collaboratively train a detection model without sharing raw data, providing privacy preservation and increased detection in more nodes. Along with Federated Learning, Explainable Artificial Intelligence (XAI) has been used to improve deep-learning cybersecurity services and the confidence in the services. This survey provides a complete analysis of the combination of federated learning and explainable hybrid deep learning systems used in the detection of cyber-attacks. This is research of the intrusion detection systems based on FL, explainable artificial intelligence systems, privacy-preserving systems, and the machine learning and deep learning systems used in the cybersecurity field. In addition, the challenges of data heterogeneity, costs of communication, privacy leakage, and the explainability gap are examined. The last portion of this survey covers the unfulfilled research needs and the paths this uncharted research can take. This can foster the development of systems to detect cyber-attacks that are explainable, private, and safe in next-generation distributed systems.

Keywords: Federated Learning, Explainable Artificial Intelligence, Intrusion Detection System, Cybersecurity, Deep Learning, Big Data Analytics.

Introduction

Developments and advancements of cloud computing, IoT, and big data analytics have resulted in sustained and evolving attacks on networks and systems. Traditional IDS use a central structure, which creates problems related to the overhead of communication, privacy, and scalability.

Federated Learning (FL) addresses issues of privacy and security. FL provides a method of distributed learning that can build models without the need to share the data. Several studies cite the ability of federated learning to build intrusion detection models in distributed systems [1]. The use and application of federated learning in cybersecurity has proven the framework's ability to create scalable, privacy-centric systems for intrusion detection [2]. Privacy-preserving federated learning has been explored to counter security concerns in the framework of model training [3].

In addition, the use of explainable Artificial Intelligence (XAI) has become prominent to address security systems' needs for deeper understanding and greater trust related to the use of deep learning. The

explainable AI approach has become popular to improve decision making and model trust in cybersecurity [4]. Explainability challenges and opportunities in the domain of intrusion detection have largely been examined to elevate the trustworthiness of intelligent security [5].

Exploration of the combination of federated learning and explainable AI is still in the early stages; the combination of these with hybrid deep learning for the detection of cyber-attacks is virtually unexplored. Thus, this survey intends to address and classify the current state to establish a basis for the development of large scale, privacy-preserving and explainable cyber-attack detection systems in distributed big data systems.

Background Concepts

Federated learning is a union-based machine learning scheme. Here, multiple participants (be it devices or institutions) can work collaboratively in unison without the need to reveal their personal, private, or raw data to the rest. During this process, participants create and feed local models, the parameters of which (also referred to as model updates) are collected by the central server and are used to formulate a global model, which is distributed back to the participants. Within a federated learning framework, we can distinguish multiple local models, a central server, and a globally coordinated model. For instance, the central server uses the Federated Averaging (FedAvg) algorithm to combine and disseminate model updates to all local participants (clients) [2]. Since raw and local data remain unshared, federated learning maintains privacy. This has significant advantages and applications within the context of many cybersecurity frameworks, in particular, in distributed and cyber-physical or IoT environments [3].

The field of Explainable Artificial Intelligence (XAI) is focused on making ML and DL models more interpretable. In the field of cybersecurity, many commonly utilized models are essentially black boxes and do not support behavior predictions [7]. The introduction of XAI can provide model interpretation to a least a degree to security analysts. This is more likely to result in a more informed and trusted decision-making process in particular, the cybersecurity realm [4]. Other examples of XAI are SHAP, LIME, and feature importance. These are particularly useful in making models interpretable in the mid-to-late stages of attack behavior prediction and outline the rationale of a particular cybersecurity decision [5].

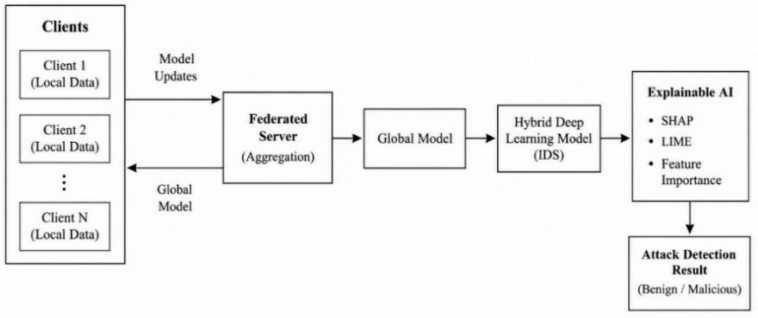

Figure.1. Federated Learning and Explainable AI Framework

Federated Learning for Cyber-Attack Detection

Federated Learning (FL) facilitates the use of distributed intrusion detection systems in privacy-sensitive environments because it does not require the transfer of raw data. Instead, models are trained on the local devices, and only the update to the model is sent to the central server.

FL has been effective in the protection of vehicular cyber-physical systems, distributed systems, and systems of the Internet of Things (IoT). In the systems of interest FL can improve privacy, and create a scalable system to perform joint attack detection [1]. In the privacy-preserving system of federated learning, security is provided in the process of model training [3]. In the case of the networks of IoT, federated learning systems have the capability of intrusion detection while the sensitive data remain in the user devices [6].

Table 1. Comparison of Existing FL-Based IDS Studies

Author	Application Domain	FL Technique	Advantages	Limitations
Safavat and Rawat [1]	Vehicular CPS	Asynchronous FL	Scalable learning	Limited evaluation
Buyuktanir et al. [2]	Distributed IDS	Federated Learning	Privacy preservation	Survey-based study
Bunko et al. [3]	Privacy-Preserving IDS	Privacy-Preserving FL	Enhanced security	Computational overhead
Khraisat et al. [6]	IoT IDS	Federated Learning	Data confidentiality	Resource constraints

Explainable Hybrid Deep Learning for Cybersecurity

The popularity of deep learning models in intrusion detection systems can be attributed to their sophisticated detection mechanisms [9]. Hybrid deep learning models fuse different deep learning architectures to enhance the classification of different attack types. These models are not completely transparent. Explainable Artificial Intelligence (XAI) aims at model transparency and provides justification for model predictions [4]. Predictive transparency is especially critical in trust engineering and security analysis [5]. In the context of attack detection, prominent methods of the XAI approach rely on SHAP, LIME, attention, and feature importance, among others [7].

Table 2. Comparison of XAI Techniques

Technique	Interpretability	Advantages	Limitations
SHAP	High	Feature contribution analysis	Computationally expensive
LIME	High	Local explanations	May be unstable
Attention Mechanisms	Medium	Highlights important features	Model-dependent
Feature Importance	Medium	Easy to understand	Limited interpretability

Comparative Analysis

Federated Learning (FL) and Explainable AI (XAI) provide innovative approaches regarding privacy and model explainability. While the promotion of FL and XAI as separate domains has been considerable, research at the intersection of FL and XAI has mostly been absent. Communication costs, privacy issues, and the lack of real-time model implementation/invocation, among other things, still pose a challenge in the domain.

Table 3. Comparative Analysis of Existing Studies

Ref.	Federated Learning	Explainability	Distributed Environment	IDS Application	Research Gap
[1]	Yes	No	Yes	Yes	Lack of explainability

[2]	Yes	No	Yes	Yes	Limited deployment
[3]	Yes	No	Yes	Yes	Communication overhead
[4]	No	Yes	No	Partial	No FL integration
[5]	No	Yes	No	Yes	Limited implementation
[6]	Yes	No	Yes	Yes	Resource constraints
[7]	No	Yes	No	No	Generic XAI focus
[8]	Yes	No	Yes	No	Limited IDS focus
[9]	No	No	No	Yes	Privacy concerns
[10]	Yes	No	Yes	No	Privacy leakage risks

Research Gaps

- Combination of Federated Learning and Explainable AI.
- Intrusion detection systems and their use of transformers.
- Federated model aggregation communication costs.
- Privacy of model updates.
- Equitable validation of benchmarking datasets.
- Its implementation and assessment in the field, and in real time.

Future Research Directions

Improved methods of Federated Learning (FL), Explainable Artificial Intelligence (XAI), and emerging deep learning driven Intrusion Detection Systems (IDS), are having an impact but are still facing difficulties with their implementation. Future research will have the potential to identify the equilibrium between the privacy, explainability and the scalability of systems while also accounting for the rapid threat detection. The following will drive the evolution of more advanced and reliable frameworks within cyber security.

Explainable Federated Learning: Privacy and explainability within Intrusion Detection Systems (IDS) need to be addressed by Federated Learning with XAI. This will help security analysts to detect certain attacks even if data remains in a secured location.

Transformer-based IDS: The IDS of the future will leverage the ability of transformer-based architectures to recognize complex long-range patterns and cyber threats to facilitate advanced and powerful detection mechanisms within systems.

Privacy-preserving XAI: There is an immediate need for the development of privacy-preserving XAI that are able to provide explanations while preserving the confidentiality of information and the security of the model.

Edge and IoT Security: The rapid growth of the Internet of Things (IoT) and edge devices is driving the demand for advanced, rapid and lightweight IDS. Future IDS based on FL need to ensure a continuous and real-time security that is scalable and elastic within a constrained space.

Autonomous Cyber Defense: Cybersecurity systems in the future will not be focused primarily on attacks. A key focus will be on systems that respond without the need for human involvement. Cyber Defense Systems that identify cyber threats and that will explain and respond to the threats in an autonomous manner will enhance the robustness of the systems being defended.

Conclusion

Federated Learning (FL) and Explainable AI (XAI) offer viable frameworks to develop cyber-attack detection systems in distributed big data contexts. FL focuses on training models in collaboration while

preserving users' data and increasing efficient scalability across distributed systems. On the other hand, XAI improves the systems' trust and transparency in detecting an intrusion through the interpretable explanations of model predictions. Additionally, hybrid deep learning models improve the detection of attacks by learning the complicated patterns in the network's traffic. However, the combination of FL and XAI is a developing research area due to the combination of challenges such as communication overhead and privacy and interpretability of models. The next-generation frameworks for cybersecurity systems must combine privacy-preserving federated learning and explainable deep learning and push advanced deep learning frameworks. The objective is to design cyber-attack detection frameworks that are scalable, flexible, and intelligent to meet the requirements of distributed environments.

References

1. S. Safavat and D. B. Rawat, "Asynchronous Federated Learning for Intrusion Detection in Vehicular Cyber-Physical Systems," *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Hoboken, NJ, USA, 2023, pp. 1-6, doi: 10.1109/INFOCOMWKSHPS57453.2023.10225917.
2. Buyuktanir, B., Altinkaya, Ş., Karatas Baydogmus, G. *et al.* Federated learning in intrusion detection: advancements, applications, and future directions. *Cluster Comput* **28**, 473 (2025). https://doi.org/10.1007/s10586-025-05325-w
3. Bunko, T., Johnstone, M.N., Yang, W. *et al.* A survey of privacy-preserving federated learning for intrusion detection systems. *Artif Intell Rev* **59**, 125 (2026). https://doi.org/10.1007/s10462-026-11519-4
4. Charmet, F., Tanuwidjaja, H.C., Ayoubi, S. *et al.* Explainable artificial intelligence for cybersecurity: a literature survey. *Ann. Telecommun.* **77**, 789–812 (2022). https://doi.org/10.1007/s12243-022-00926-7
5. Pawlicki, M., Pawlicka, A., Kozik, R. *et al.* The survey on the dual nature of xAI challenges in intrusion detection and their potential for AI innovation. *Artif Intell Rev* **57**, 330 (2024). https://doi.org/10.1007/s10462-024-10972-3
6. Khraisat, A., Alazab, A., Alazab, M. *et al.* Federated learning for intrusion detection in IoT environments: a privacy-preserving strategy. *Discov Internet Things* **5**, 72 (2025). https://doi.org/10.1007/s43926-025-00169-7
7. A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," in *IEEE Access*, vol. 6, pp. 52138-52160, 2018, doi: 10.1109/ACCESS.2018.2870052.
8. Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. 2019. Federated Machine Learning: Concept and Applications. ACM Trans. Intell. Syst. Technol. 10, 2, Article 12 (March 2019), 19 pages. https://doi.org/10.1145/3298981
9. Ahmad Z, Shahid Khan A, Wai Shiang C, Abdullah J, Ahmad F. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Trans Emerging Tel Tech. 2021;32:e4150. https://doi.org/10.1002/ett.4150
10. Viraaji Mothukuri, Reza M. Parizi, Seyedamin Pouriyeh, Yan Huang, Ali Dehghantanha, Gautam Srivastava, "A survey on security and privacy of federated learning", Future Generation Computer Systems, Volume 115, 2021, Pages 619-640, ISSN 0167-739X, https://doi.org/10.1016/j.future.2020.10.007.