Pegem Journal of Education and Instruction, Vol. 15, No. 4 , 2025 (pp. 5198-5206)

RESEARCH ARTICLE WWW.PEGEGOG.NET

Trustworthy AI Federated Learning Frameworks for Secure IoT Applications

Anitha Gurram*[1] , Dr. Gaurav Tyagi[2]

1. Research Scholar, Dept. Of Computer Science & Engineering, Chaudhary Charan Singh University, Meerut, Uttar Pradesh, India, 250005. email: ganitha29685@gmail.com

2. Associate Professor, Dept. Of Computer Science & Engineering, Chaudhary Charan Singh University, Meerut, Uttar Pradesh, India, 250005.

Abstract-

This paper proposes a novel framework that unifies GDPR-compliant federated learning architectures with trust-based malicious node identification for IoT networks. By combining privacy-preserving mechanisms with trust scoring, the framework ensures both legal compliance and resilience against adversarial participants. The proposed system leverages AI-driven federated learning enhanced with Zero-Knowledge Proofs (ZKPs) to validate model updates without exposing sensitive data. Experimental validation on IoT datasets demonstrates improved robustness, scalability, and compliance compared with existing FL approaches.

Keywords: Deep federated learning (DFL),, Artificial Intelligence, Internet of Things, Zero-Knowledge Proofs, general data protection regulation (GDPR),Trust-Based Security·

I. Introduction

In DFL, each IoT device retains its data locally and trains a model on it, then aggregates model updates with those of other devices to update a global model. This ensures that critical information stays on the device, lowering the possibility of data leaks and enabling the creation of models trained on a more diverse and comprehensive dataset. By using DFL in IoT, organizations can unlock the full potential of their data while maintaining the privacy and security of their users[1][2]. Therefore, creating novel AI approaches is crucial for achieving effective and privacy-enhancing smart IoT networks and applications. DFL can deal with the challenges of training ML models on decentralized, sensitive data while preserving privacy. DFL can provide several significant advantages for IoT applications, including the following:

• Privacy protection: By allowing models to be trained on decentralized data, DFL ensures that sensitive information remains on individual devices and is never transmitted to a centralized server. This protects users' privacy and reduces the risk of data breaches.

• Improved model performance: DFL enables the training of ML models on a more diverse and comprehensive dataset, as each IoT device contributes its local data. This results in improved model performance and accuracy compared to models trained on centralized data alone.

• Reduced latency and bandwidth requirements: DFL reduces the amount of data transmitted between devices and a centralized server, as only model updates are transmitted. This reduces latency and bandwidth requirements, making it well-suited for low-power,resource-constrained IoT devices.

• Scalability: DFL enables ML models to be trained on a large number of IoT devices, allowing organizations to scale their systems as needed.

• Offline training: DFL allows for training to occur even when devices are offline, making it well-suited for IoT devices with limited or inconsistent connectivity.

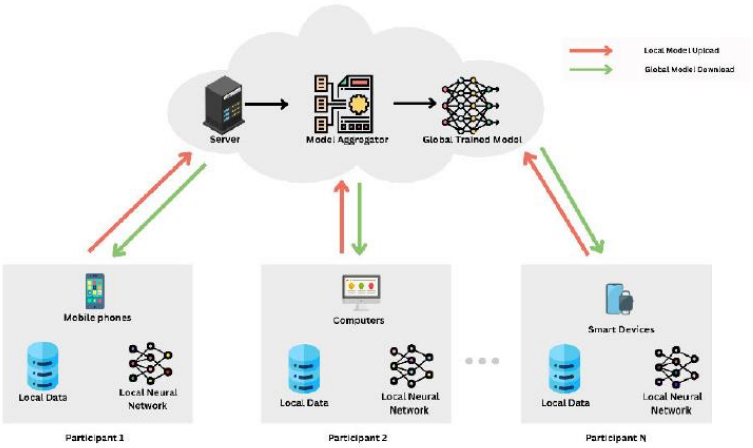

Figure 1. An architecture diagram of Deep federated learning.

DFL is an exciting field of ML that allows distributed training of a shared model without requiring the raw data to be shared. In the next section, we will explore the different types of DFL that can be used depending on the nature of the data and the goals of the training[3][4]. By understanding these different approaches, we can choose the most appropriate. DFL technique for our specific needs and applications.

II. RELATED WORK

DFL stands at the forefront of technological innovation, particularly in the realm of Internet of Things (IoT) services. This cutting-edge approach to machine learning is designed to address the challenges posed by decentralized and distributed IoT environments. In Table 1, DFL for IoT services is summarized. The architecture [5] accomplished collaborative fairness via local dependability, participation level, and transaction points, and the concept of different datasets is incorporated for privacy throughout the review process [6]. The proposed technique, as per experimental data, proved resilient against poisoning attacks, leveraging networked devices for malware defense and threat classification. Based on the findings of [7], the authors of the paper proposed a distributed DL architecture called Fair and Privacy-Preserving Federated Deep Models (FPPDL). By introducing the concept of geographical credibility and transaction points, and using blockchain technology for decentralization, this method seemed to improve impartiality in cooperative DL. The use of encryption in tandem with Partially Private GAN helped the system achieve its dual goals of confidentiality and precision. The MNIST accuracy achieved by the centralized framework using CNN architecture on P4 (four parties involved in the experiment) is 96.58%[8].

The susceptibility of DNNs to white-box inference attacks in both centralized and FL settings was proposed. The paper proposed novel algorithms tailored to the white-box setting to exploit privacy vulnerabilities of the stochastic gradient descent algorithm used in DNN training[9][10].

The paper evaluated the efficacy of white-box membership inference attacks against DL models and demonstrated the susceptibility of even well-generalized models to such attacks. It also showed how adversarial participants in FL can successfully run active membership inference attacks against other participants [11].

Table 1: Survey table of services and limitations

Services	Techniques used	Contribution	Limitations	Accuracy
Wormhole attack Monitoring	Cascaded FDL	Data security and privacy aspects computation and processing issues are addressed	Lack of CNT	97%
Reputation-based Mechanism	Dynamic Asynchronous Anti-poisoning	Superior performance, Reputation Aware, Communication Reduction, Training time reduced by 30%	No real-world Evaluation, Single Dataset Evaluation, and Limited vulnerability coverage	Not provided
Collaborative and Industrial Cyber-physical system Intrusion Detection	Federated Deep Learning	Improved detection accuracy, Preserved privacy and collaborative approach	Limited to homogeneous network and supervised learning tasks, Required a trusted aggregator and communication overhead	99.27 ± 0.79%, 99.20% (k=7)
Zero-day Botnet attack detection	Federated Deep Learning	Improved detection accuracy	Limited Model Capacity	99.79 ± 0.01%
Industrial Control System and FL	LSTM and CNN	Developed a novel DFL-Based ICS attack detection method, preserved privacy and security	Height computational cost and data privacy concerns	90.83%
Cyber-Physical System	Transformational Approach	Architectural Practices and System Integration	Interoperability issue, Scalability concerns	99.20%
Heating Load Demand Forecasting	Secure Federated Deep Learning-based Approach	Server has the lowest forecasting error, Global supermodel able to predict the heating load demand with correlation coefficient of 98.00%, 93.00%, and 70.00%.	Limited Dataset Availability	99.00%.
UAV-assisted RIS	LSTM and FL	Demonstrate the effectiveness with UAVs and RIS, Proposed a Novel Resource Allocation Strategy	Assumes a fix communication topology, Uses a simplified channel model	Not provided

The paper [12] provided a DFL approach to enable secure and private Point of Interest (POI) management in Cyber-Physical Systems (CPS). The recommended technique was tested using two real-world datasets, and the results showed promise in terms of achieving the design's goals.The other way is to explore methods to reduce the computational complexity of DL algorithms while making them more understandable. The effectiveness of FDL techniques in enhancing individual privacy in the IoT is contrasted in. FL systems are contrasted with blockchain, intrusion/malware tracking systems, and some other IoT application types. Additionally, the authors discussed the potential security and privacy issues with FL-based systems, in an experimental study of three DL models, three IoT connectivity datasets were employed. The results showed that FL algorithms provided superior confidentiality for data from IoT devices and achieved greater accuracy in spotting risks than centralized ML approaches.

III. Research Challenges: Deep Federated Learning

DFL is a cutting-edge approach to machine learning that has garnered attention for its potential to address challenges FIGURE 4. Research challenges in the field of Deep federated learning.in distributed learning while prioritizing data privacy.In a decentralized learning environment, preserving sensitive information and preventing privacy violations become paramount concerns. In the following sections, we delve into some of the key research challenges associated with DFL.

Figure 2. Research Challenges in the Field of Deep Federated Learning.

The subsequent discussion highlights the complexity of these challenges and emphasizes the need for continuous research and development in these areas. By doing so, we can ensure that DFL remains a secure, compliant, and ethical approach to machine learning, paving the way for its continued adoption and application in diverse industries. Figure 4 illustrates the research challenges, providing a visual representation of the multifaceted landscape that researchers and practitioners must navigate to enhance the effectiveness and robustness of DFL.

IV. PROPOSED FedTrust METHODOLOGY

The proposed FedTrust approach utilizes the concept of federated learning and communities to efficiently maintain trustworthiness. The proposed approach trained the model using edge nodes with a modified federated learning implementation architecture. This section will elaborate on the proposed approach architecture, along with dataset features, splitting, and the training process of deep federated learning. Federated learning is used in our approach, which frees edge nodes from sharing

data as they train their models. Unlike with some other available options, this one guarantees that no private data will be disclosed.

The proposed solution is well-suited to dealing with large-scale IoT networks because to deep federated learning's ability to distribute the computational burden across edge nodes. This saves a lot of time and money compared to traditional, centralized educational models. Defense against external forces: Our novel trust dataset with trust parameters makes the proposed approach more resilient to common attacks like whitewashing and badmouthing. We will demonstrate how FedTrust performs in comparison to other solutions under various forms of attack.

Thanks to its ability to continuously learn and adapt to the changing behavior of IoT nodes, our technique is particularly well-suited for dynamic and heterogeneous IoT networks[13][14].

Table 2. Analysis of Current State-of-the-Art Methods

Approach	Technique	Application
Decentralized Trust Management	Blockchain-empowered Federated Learning	IoT Security
Trust-driven Reinforcement Selection	Federated Learning	IoT Devices
Application or Infrastructure Co-design	Real-time Edge Video Analytics	IoT Analytics
Trust-augmented Deep Reinforcement Learning	Federated Learning Client Selection	IoT Networks
Hierarchical Blockchain-based FL	Collaborative IoT Intrusion Detection	IoT Security
Trusted Feature Aggregator FL	Malicious Attack Detection	IoT Security
Blockchain-supported FL	Securing Critical IoT Infrastructures	IoT Security

The proposed approach is based on the modified implementation of federated learning by employing a domain server that handles the communities and trains the model by dividing the dataset. The proposed approach architecture consists of four major components, i.e., Global dataset, global model. The global dataset is split into several parts based on the available domain servers and delivered to each domain server[15][16]. server. Further, the domain server selected the nodes from the communities based on their capabilities and competency to train the received global dataset. After that, the domain server further split the dataset into the number of selected nodes, and each part to the specific nodes for training purposes.

V. EXPERIMENTAL ANALYSIS

We propose an ensemble learning approach for trust-based intrusion detection in IoT environmentss, using knowledge, reputation, and experience as trust management components. Our approach is compared with two existing approaches, PoTC [17] and DDQN-Trust, for a comprehensive evaluation of its performance. The dataset consists of 19 trust features, and we employ an ANN as the base model. The parameters used to build the dataset were both randomly chosen and pre-established. The dataset was produced by doing the following steps:

The number of nodes and edges, as well as the network's overall size, were created at random.

• Each node was assigned an arbitrary value based on its knowledge, reputation, and experience.

• The weights for each metric were assigned random values within the range of 0 to 1.

• All of the factors that make up the knowledge metric how credible, accurate, reliable, compliant, capable, available, and responsive-were given arbitrary values between 0 and 1 to make up a single score.

• Parameters of nodes, such as their ratings, user engagement, and responsiveness, as well as the number of data breaches and security vulnerabilities, were given arbitrary values between 0 and 1 for the reputation measure.

• Random values between 0 and 1 were assigned to the parameters of interaction frequency and transaction success rate, transaction time, communication quality, and resource utilization, data sharing behavior, and cooperation level, and end-to-end packet delivery to calculate the experience metric.

• To ensure that all parameters were between 0 and 1, the dataset was normalized using min-max normalization method.

• Finally, the dataset was split into training and testing sets with a 70:30 ratio.

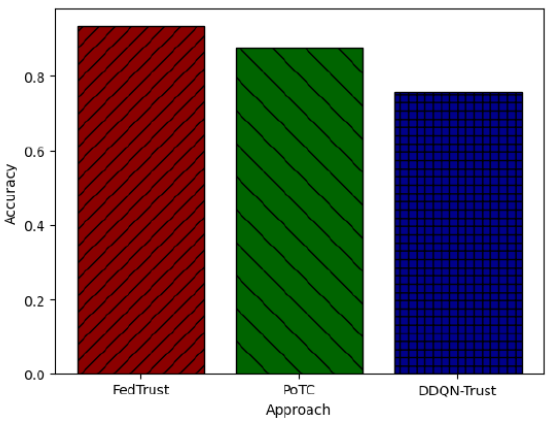

Figure 3.A.

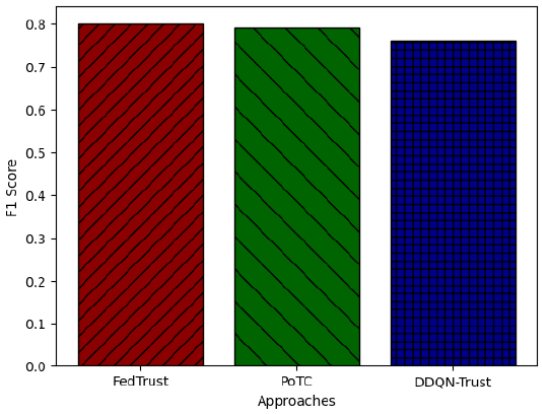

Figure 3.B.

3.(a) Accuracy comparison of FedTrust with existing approaches

3.(b). F1 score comparative analysis of FedTrust with existing

The evaluation of the proposed model has been performed using the accuracy metric. Accuracy is defined as the ratio of the correctly classified data points to the total number of data points in the dataset. The accuracy in the proposed model has been created, as illustrated:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP (True Positive) is the number of correctly classified malicious nodes, TN (True Negative) is the number of correctly classified benign nodes, FP (False Positive) is the number of benign nodes classified as malicious, and FN (False Negative) is the number of malicious nodes classified as benign[18][19].

The F1 Score is a measure of a model's accuracy, calculated as the harmonic mean of the model's precision and recall. It is a commonly used metric to evaluate the performance of machine learning models, particularly in the context of binary classification problems. In our study, we used the F1 score to evaluate the performance of our proposed ensemble learning approach, FedTrust, and compared it with two other existing approaches: PoTC and DDQN-Trust.

Precision is a performance metric that calculates the proportion of true positive instances among the total instances predicted. as positive[20][21]. It is computed as the ratio of true positives to the sum of true positives and false positives[22]. Precision is an important evaluation metric, especially in scenarios where false positives can cause significant damage or false alarms.

VI. CONCLUSION

A novel ensemble learning approach for detecting malicious nodes in an IoT environment using trust management components such as knowledge, reputation, and experience. The proposed approach utilizes an ANN as a base model to classify the nodes as malicious or benign. To optimize the model's performance of the model, we will use Keras Tuner to search for the optimal hyperparameters of the ANN, such as the number of hidden layers, the number of neurons in each layer, the activation function, the optimizer, and the learning rate. The proposed architecture consists of three main components: the data acquisition module, the trust management module, and the decision-making module. The proposed approach can be further extended to investigate scalability and robustness of the proposed approach in real-world scenarios with a large number of nodes and complex IoT architectures.

DFL has emerged as a promising distributed AI technique that can enable private and scalable IoT services and applications. The article provides a comprehensive overview of DFL and various DFL services and applications. The challenges to the privacy of individual users and devices are also identified, which can have legal as well as ethical implications. DFL-based systems need to prioritize enhancing security and privacy protections to safeguard sensitive data during collaborative learning. Keeping this in mind, we presented the initial idea of a GDPR-Compliant DFL architecture in this paper. This architecture combines various privacy-preservation techniques employed by existing

DFL-based systems studied in the literature. We intend to further extend this idea into a comprehensive framework that will be formally verified and evaluated in our future work.

REFERENCES

[1] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on Internet of Things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1125–1142, Oct. 2017.

[2] H. Lin, K. Kaur, X. Wang, G. Kaddoum, J. Hu, and M. M. Hassan, "Privacy-aware access control in IoT-enabled healthcare: A federated deep learning approach," *IEEE Internet Things J.*, vol. 10, no. 4, pp. 2893–2902, Feb. 2023.

[3] Y. Liu, J. Wang, Z. Yan, Z. Wan, and R. Jäntti, "A survey on blockchain-based trust management for Internet of Things," *IEEE Internet Things J.*, vol. 10, no. 7, pp. 5898–5922, Apr. 2023.

[4] P. Ravichandran, C. Saravanakumar, J. D. Rose, M. Vijayakumar, and V. M. Lakshmi, "Efficient multilevel federated compressed reinforcement learning of smart homes using deep learning methods," in *Proc. Int. Conf. Innov. Comput., Intell. Commun. Smart Electr. Syst. (ICSES)*, Sep. 2021, pp. 1–11.

[5] B. Yin, H. Yin, Y. Wu, and Z. Jiang, "FDC: A secure federated deep learning mechanism for data collaborations in the Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6348–6359, Jul. 2020.

[6] S. Kaspour and A. Yassine, "A federated learning model with short sequence to point mechanism for smart home energy disaggregation," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2022, pp. 1–6.

[7] H. Mohamed, N. Koroniotis, and N. Moustafa, "Digital forensics based on federated learning in IoT environment," in *Proc. Australas. Comput. Sci. Week*, Jan. 2023, pp. 92–101.

[8] J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, and W. Zhang, "A survey on federated learning: Challenges and applications," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 2, pp. 513–535, Feb. 2023.

[9] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for Internet of Things: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1622–1658, 3rd Quart., 2021.

[10] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100379.

[11] K. Lakshmanna, R. Kaluri, N. Gundluru, Z. S. Alzamil, D. S. Rajput, A. A. Khan, M. A. Haq, and A. Alhussen, "A review on deep learning techniques for IoT data," *Electronics*, vol. 11, no. 10, p. 1604, May 2022.

[12] S. H. Shah and I. Yaqoob, "A survey: Internet of Things (IoT) technologies, applications and challenges," in *Proc. IEEE Smart Energy Grid Eng. (SEGE)*, Aug. 2016, pp. 381–385.

[13] J. A. Josephine, S. Senthilkumar, R. Rajkumar, and A. Kumar, “Detection of authorized nodes to provide an optimal secure communication in amalgamated internet MANET,” in *Proc. Int. Conf. Internet Things*. Berlin, Germany: Springer, 2023, pp. 93–102.

[14] A. Farraj, “Coordinated security measures for industrial IoT against eavesdropping,” in *Proc. IEEE Texas Power Energy Conf. (TPEC)*, Feb. 2023, pp. 1–5.

[15] P. D. Rosero-Montalvo, Z. István, P. Tözün, and W. Hernandez, “Hybrid anomaly detection model on trusted IoT devices,” *IEEE Internet Things J.*, vol. 10, no. 12, pp. 10959–10969, Jun. 2023.

[16] P. Benlloch-Caballero, Q. Wang, and J. M. A. Calero, “Distributed dual-layer autonomous closed loops for self-protection of 5G/6G IoT networks from distributed denial of service attacks,” *Comput. Netw.*, vol. 222, Feb. 2023, Art. no. 109526.

[17] M. Abbasi, M. Plaza-Hernández, and Y. Mezquita, “Security of IoT application layer: Requirements, threats, and solutions,” in *Proc. 13th Int. Symp. Ambient Intell.* Berlin, Germany: Springer, 2023, pp. 86–100.

[18] R. Verma and S. Chandra, “RepuTE: A soft voting ensemble learning framework for reputation-based attack detection in fog-IoT milieu,” *Eng. Appl. Artif. Intell.*, vol. 118, Feb. 2023, Art. no. 105670.

[19] S. Subramani, M. Selvi, A. Kannan, and S. K. Svn, “Review of security methods based on classical cryptography and quantum cryptography,” *Cybern. Syst.*, pp. 1–19, Jan. 2023.

[20] L. Ouyang, F. Wang, Y. Tian, X. Jia, H. Qi, and G. Wang, “Artificial identification: A novel privacy framework for federated learning based on blockchain,” *IEEE Trans. Computat. Social Syst.*, early access, Feb. 1, 2023, doi: 10.1109/TCSS.2022.3226861.

[21] A. O. Philip and R. K. Saravanaguru, “Multisource traffic incident reporting and evidence management in Internet of Vehicles using machine learning and blockchain,” *Eng. Appl. Artif. Intell.*, vol. 117, Jan. 2023, Art. no. 105630.

[22] P. K. Laboso, A. Martin, and P. Thiyagarajan, “Blockchain technologies in data science: Challenges and benefits,” in *Proc. Int. Conf. Sustain. Comput. Data Commun. Syst. (ICSCDS)*, Mar. 2023, pp. 1323–1328.