



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: <https://doi.org/10.22214/ijraset.2024.65079>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Federated Learning-Based Authentication and Trust Scoring System for Cloud IoT Security

Jyoti Rani¹, Rushan Gupta², Sameer Kumar Singh³, Manik Sood⁴, Kuldeep Kumar Chauhan⁵

Chandigarh University Mohali, India

Abstract: A number of IoT-related security challenges are faced due to the introduction of IoT devices in almost all sections, including smart homes and industrial applications. New security solutions will be required in IoT networks to tackle issues pertaining to privacy, scalability, and efficiency with solid robustness against potential vulnerabilities. Even though some occasions these are effective centralized security frameworks suffer from significant limitations, including single points of failure, increased latency, and risks associated with data centralization [1][2]. These problems worsen as IoT networks grow in size and complexity.

The below is the Federated Learning-Based Authentication and Trust Scoring System proposal toward overcoming these challenges while keeping in view the benefits offered by FL. It enables IoT devices to jointly train a global model for anomaly detection without raw data sharing [3]. This happens in a distributed manner so that data privacy is ensured, with scalability and network performance improving [4]. With this, there is a dynamic trust scoring mechanism that evaluates each device's reliability in accordance with the behavioral patterns and history of interactions [5][6]. When combining FL with the trust scoring mechanism, the proposed solution would allow an IoT network to be autonomously able to identify anomalies and manage security.

This system has shown practical applicability in such experiments in both the smart home and industrial environments to enhance IoT security, retain scalability, and maintain high privacy standards [7]. The adaptability of the system coupled with federated learning makes it apt for a variety of IoT ecosystems [8].

Keywords: Federated learning, IoT security, trust scoring, anomaly detection, distributed systems, privacy preservation, scalability, smart home security, industrial IoT, decentralized authentication.

I. INTRODUCTION

IoT has revolutionized multiple domains with advanced connectivity and the capability to process real-time data. Internet of Things devices, in one sense, have become an integral part of the modern system—from smart home devices to industrial sensors. However, such massive adoption of these devices dictates significant security challenges. Centralized security systems that aggregate and process data at a central server often suffer due to high latency, violation of privacy, and susceptibility to single points of failure [10, 11].

Centralized systems come with the risks of data breaches and offer lower scalability since they store large amounts of data. Relaying raw data to central servers involves exposure to sensitive information and added security risks [12]. In addition, the traditional anomaly detection methods may not scale to the volume of data and complexity of devices in highly wide IoT networks [13].

The rising challenges in sensor networks and machine learning have prompted the introduction of federated learning. FL empowers distributed training of machine learning models by allowing devices to train models locally on their data, sharing model updates only with a central server [3]. Therefore, with respect to privacy, raw data remains on the device, while scalability problems are decreased through decentralization of computational workloads. However, the integration of FL in IoT security systems is still a relatively unresearched domain.

This paper introduces a new Federated Learning-Based Authentication and Trust Scoring System that focuses on leveraging FL to improve the security of IoT devices. The system decentralizes anomaly detection and offers dynamic trust scoring to enhance network security while guaranteeing user privacy and the scalability of the system.

II. LITERATURE REVIEW

The massive deployment of billions of connected devices worldwide has dramatically changed the security landscape in IoT networks. Because such devices would collect and share sensitive data, there is a pressing need for robust security mechanisms to guarantee privacy, trust, and resilience against malicious attacks.

Existing successful models for IoT security, which have served well in constrained, controlled environments, suffer from scalability, privacy, and latency when deployed in large heterogeneous IoT systems.

- 1) *Centralized Security Systems for IoT*: Traditionally, these security frameworks for IoT have been based on centralized models, where all the data coming from all the devices in the network is accumulated at a central server in order to be processed. Here, machine learning algorithms are applied in order to detect an anomaly and enforce security policies.

Centralized systems are cumbersome and not scalable in large-scale IoT deployments, especially concerning potential privacy risks, latency, and bottlenecks on the performance front. As the IoT networks are growing, large central models have to process huge amounts of data, thus degrading anomaly detection and overloading central resources. In addition, data centralization goes on to introduce the vulnerabilities of privacy because raw data from all the devices is held and processed in one place; thus, it increases the possibility of hacking and unauthorized access [11][12]. As Gubbi et al. [14] pointed out, latency increases exponentially with the increase in the number of connected devices and rises exponentially to greater levels in timecritical applications such as IIoT systems and healthcare. Centralized systems pose a single point of failure with regard to cyberattack and threaten the entire network. Hence, more decentralized security architecture has to be implemented to deploy data processing across devices to minimize such risks.

- 2) *Federated Learning-A Decentralized Approach*: FL offers an exciting solution to the scalability challenges posed by centralized systems as well as to privacy challenges. Proposed first in [3] by McMahan et al., FL allows multiple devices to collaborate with each other in training an accurate global machine learning model, even without communicating raw data with each other but by exchanging only local model updates with a central server, namely, gradient information. It aggregates these updates into improving the global model. The architecture preserves device data privacy while reducing latency and bandwidth requirements associated with central transfer of data.

Since the invention of FL, it has been applied extensively in applications like mobile networks and autonomous vehicle applications. The deployment in mobile devices was demonstrated by Hard et al. [15], indicating that FL improves predictive models without invading privacy while using it. In the specific case, FL is appropriate for applications involving IoT because of the decentralized IoT network architecture. However, its application to IoT security is still relatively unexplored; however, some limited research has been carried out concerning the feasibility of introducing FL into authentication and trust mechanisms.

The largest challenge associated with applying FL to IoT systems is the heterogeneity of devices and the data generated from the same. The hardware capabilities, software configurations, and communication protocols differ from one device to another. Because of this, the resultant data is usually non-IID and poses a highly big problem for federated learning models. Zhao et al. [7] mentioned this challenge and considered the need for robust aggregation algorithms and learning models from instances on the IoT environment related to varying data distributions.

- 3) *Trust Scoring Systems in IoT Networks*: The primary role of trust scoring mechanisms in IoT security is to assess the credibility of any device based on its behavior and relationship with the other nodes within the network. Conventional trust models count on a blend of factors such as device history, communication pattern, and adhesion to protocol to provide trust scores to devices [16]. The results are then routed to security policies-for instance, grant access to network resources, or flag the possibility of a rogue device for further examination. One primary benefit of a trust score is it offers a dynamic mechanism through which reliability may be assessed, and on its merit, security policies are updated to reflect these changing conditions. Based on the fact that some of them can easily detect insider threats and malicious behavior, some researchers have recently suggested trust-based models for sensor networks and peer-to-peer systems [17]. However, most such models are intended for centralized architectures in which a central server computes and maintains all the scores for the trust of devices in the network. Given that centralization is fragile on several accounts, similar to traditional IoT security models, it has scalability as well as privacy issues.

A decentralized trust management system with federated learning might be a more scalable and resilient approach to IoT security. The evaluation of trust scores and anomaly detection would be distributed, offloading the workload from central servers and further eliminating risks associated with potential central points of failure. Yang et al. introduced a blockchain-based distributed trust management system to ensure integrity in trust scores [9]. Promising approaches but quite limited in scalability, requiring much additional work toward practical usage in large-scale IoT deployments.

4) *Integrating Federated Learning with Trust Scoring:* A federated learning integrated trust score-based system represents a new way of putting forth the limitations of a traditional IoT security system. FL has been employed to decentralize anomaly detection, thus each device assessing its behavior as part of building a global security model. In contrast, the dynamic scoring mechanism of devices based on past performance and current behavior represents the evaluation process. This two-pronged design better networks by limiting the chances of privacy violation, while reducing latency and enhancing scalability.

Zhou et al. [18] have showed how federated learning may be combined along with blockchain-based trust models for mobile networks to increase security. The same concept may be applied to IoT networks for bringing in privacy-preserving scalability to a solution in the security. By the use of FL, it would enable training on global anomaly detection models as well as trust scoring for the assessment of device behavior. The system would dynamically change the security policies to allow trusted devices to participate only within that network.

In conclusion, federated learning and trust scoring have been widely studied in other domains; however, the proposed integration of the two concepts with IoT security is quite novel and very little explored so far. The proposed system addresses the concerns of decentralized, scalable, and privacy-preserving regarding IoT security.

III. PROPOSED SYSTEM

We detail the system components and architecture of the Federated Learning-Based Authentication and Trust Scoring System proposed in the following subsection. Our system will address the largest challenges of security in large-scale IoT networks using a decentralized, privacy-preserving anomaly detection approach in combination with a trust management mechanism ensuring only credible devices contribute to the network.

1) *Federated Learning Framework:* The Federated Learning Framework would be the core of our security solution; through this, IoT devices could collaborate with each other to learn an anomaly detection model without having to share raw data. FL addresses the problem of privacy inherent to centralized machine learning systems wherein all data from the devices has to be sent to a central server. Instead of that, in federated learning, each device trains a local model of its own by using information from the device's sources of data, such as sensor readings, network activity, or system logs. Local models are constructed to focus on identifying anomalous behavior, like odd patterns in network traffic, or like odd patterns in the performance of devices that may otherwise denote a security breach.

At periodic intervals, each device is updating its local model and sharing only model parameters, that is, gradients or weights, but not the data itself, with a centralized server. Central server aggregates model updates using Federated Averaging algorithm known as FedAvg [3]. The Algorithm FedAvg works on average gradients from all participating devices and thus creates a new, globally improved model. The updated world model is then sent back to all the devices on the network. This process iterates over time, updating the model but still enabling it to enjoy all varieties of data being generated from the network while keeping individual device information private.

The FL approach encompasses the following important issues:

- **Data Privacy:** Since raw data leaves at no time from the device, the system raises the barrier of breach of data substantially and safeguards sensitive information of the user by not letting it reach the central server.
- **Scalability:** The training process is decentralized; the computational load is dispersed across the devices so that it remains scalable even with increasing numbers of devices.
- **Robustness to Single-Point Failures:** Unlike the traditional centralized setup, federated learning disperses intelligence throughout the network, thus avoiding potential risks presented by a single point of failure [11][12].

In addition to that, Zhao et al. [7] also reported in their study that federated learning is particularly appropriate for heterogeneous IoT environments wherein devices could produce data that would be non-IID. Hence, this aspect makes FL adaptable to the wide variety of IoT devices into largescale deployments.

2) *Trust Scoring Mechanism:* A Trust Scoring Mechanism is proposed in the paper to further improve its security. Here, each device that would be a part of the network has been assigned a trust score, which is refreshed periodically based on behavior and other interactions, thereby assisting in monitoring the participation of potential malicious or compromised devices during the federated learning process.

Based on many scores, the trust score is computed.

- **Behavioral Consistency:** Devices are measured on whether the behavior they are producing is consistent with past behavior. A large number of significant divergencies in behavior away from expected behavior may indicate a security problem, for instance, an infected device.
- **Network Level Interaction:** Number of communications and type with all other devices within the network, as well. Those devices which send communications to suspicious or unknown entities are marked for further investigation.
- **Model Update Integrity:** Devices are rated according to how good the model updates they upload to the server. This means the devices which frequently upload meaningful updates receive trust ratings, while the ones which upload false, deceitful, or malicious updates receive low trust scores.
- **Past Security Events:** The score of trust is conscious of the past security events like anomalous behavior or network disruption. Attention is given to the devices with a history of security anomalies [16].

Trust score is the major determinant of the degree of access that a device is allowed into a network. With a higher trust score, it assigns full permission to a device. With lower trust scores, devices are banned and at times, an extra authentication check has to be done. This dynamical type of access control ensures that network critical activities will be executed by trusted devices. Therefore, attacks are limited to breach any of the security features.

- 3) *Model Aggregation and Trust-Based Response:* This collection of the local updates will subsequently be used as input in a global model developed by the FedAvg algorithm. The updated model is then transmitted to all the participant devices. This setup will continue to progress in such a way that with time, the anomaly detection capabilities of the model are enhanced, and the security provided by the entire system increases.

The trust scoring mechanism within the system generally decides how the system is going to respond to the presence of potential security threats. A high-scoring device will most definitely participate in federated learning in its entirety, which includes contributing model updates and accessing network resources. A device with a lower trust score will likely be probed further through means such as increased model updates or additional authentication layers.

For an extension, such a system could be more aggressive when it detects devices that present with suspicious behavior history or not meeting specific set thresholds in their respective trust scores. Quarantining the same devices from the rest of the network ensures safety verification before any access to any part of the network is granted. This ensures that compromised or malicious devices do not damage the integrity of the network.

This trust-based approach further prevents malicious devices from any kind of tampering in terms of sending bogus model updates into federation learning. It is assured that the weight given to the global model updates by trustworthy devices will be much higher, and therefore the global model will be stronger and accurate. This further increases the security and at the same time also increases the performance of the model about anomaly detection regarding the network [15][7].

- 4) *Advantages of the Proposed System:* The Federated Learning-Based Authentication and Trust Scoring System has the following significant advantages against conventional centralized security models:

- **Improved privacy:** Since the data is localized in the device, there's a lesser possibility of privacy breach, which is a specific concern in IoT networks where most devices collect sensitive user information [9].
- **Scalability:** Decentralized nature in federated learning assures that the system will be scalable to a great extent and allows for large numbers of devices with no overburdening a central server.
- **Attack Resiliency:** The federated learning symbiosis with scoring of trust will immediately detect and isolate rogue devices so that they cannot impact the network.
- **Dynamic security policies:** The application of trust scores enables the system to adapt in real time, based on its security policy, thus providing flexible and responsive security management.

All the scalability, privacy, and security requirements of the IoT network are met by the proposed system implementing integration of FL with trust scoring. It would be promising future research to aggregate optimization of algorithms and further refinement of the trust scoring metrics to improve performance in such a system

IV. DESIGN AND IMPLEMENTATION

Describe the experimental setup that we have for implementing the Federated Learning-Based Authentication and Trust Scoring System and discuss the tools and technologies that we use. The design is privacy-sensitive, scalable, and real-time in an attempt to address the big challenges of impacting IoT networks.

1) *IoT Device Setup and Data Collection*: The experimental setup is designed with different types of IoT devices along with mainly Raspberry Pi and ESP32 microcontroller units, where the units are installed with different sensors simulating various use cases. These sensors include temperature sensors, humidity sensors, and motion sensors chosen based on the fact that they are mostly prevalent in smart homes as well as industrial settings.

- Temperature Sensors were programmed to indicate spurious temperature peaks, which could be due to things that are unrelated to faults in equipment or tampering and which a bad employee might attempt to do.
- Humidity Sensors: The sensors sensed and monitored environmental conditions that would alert the system of any unexpected changes, such as that induced by faulty equipment or access to controlled areas.
- Motion Sensors: They detect and monitor for any unauthorized movements-for instance in security-sensitive areas of smart homes or industrial premises.

For each device, this kind of sensor processed the acquired data in real-time. A lightweight anomaly detection model was used for categorizing an anomaly, which according to the present system, is defined as a sensor behavior that falls out as an odd one. For this, SVM and K-means clustering techniques were used. Local anomaly detection aimed at identifying threats or problems in real-time; this excludes continuous cloud communication and allows a reduction in latency.

2) *Data Handling and Preprocessing*: All acquired data from sensors undergo a preprocessing phase with noise reduction techniques such as moving average smoothing, for instance, to remove irrelevant fluctuations. Then, a normalization of data is carried to ensure uniformity on devices that have different sensing ranges to result in more consistent local model performance with anomaly detection. Once processed locally, the models did detect potential anomalies and only forward model updates-i.e., learned parameters-to the central server but not the raw data itself. This serves central to the goal of privacy preservation. An important concern in many IoT applications is maintaining privacy, especially given that devices are deployed in sensitive environments such as personal homes or even critical industrial setups.

3) *Federated Learning and Trust Scoring Implementation*: The federated learning (FL) architecture was presented with both TensorFlow Federated (TFF) [1] and the widely adopted privacy-preserving machine learning frameworks PySyft [20]. TFF library supports secure, distributed training of machine learning models on IoT devices. On the other hand, PySyft provides secure and privacy-focused operations such as techniques of differential privacy and secure multi-party computation (SMPC). These devices trained local models using data coming from sensors. They could classify anomalies directly. Periodically, they would send updates of their respective models, or gradients, to the central server. This central server uses AWS Lambda [21], which aggregates updates by using the Federated Averaging algorithm [3]. Using AWS Lambda ensured efficient serverless computing and enabled dynamic scaling of the central server according to how many IoT devices connect to it.

4) *Trust Scoring Mechanism*: A trust scoring mechanism was developed in Python, implemented within the FL framework. Each device is assigned a score based on multiple factors:

- Pattern of Behavior: It regularly observed expected patterns of behavior over time. Any sudden change from the norm indicated a possible security problem, which lowered the trust score.
- Model Update Quality: Devices offering meaningful and accurate model updates were trusted to a large extent. It penalized malicious and unreliable devices trying to give misleading or erroneous updates.
- Communication Interactions: Devices that displayed suspicious or abnormal communication behavior, such as frequent connections with unknown or unauthorized devices had lower scores for trust.

The trusted scores were allowed access to sensitive network resources and full participation in federated learning when devices had more trust scores. Conversely, those with lesser trust scores were either quarantined or restricted from accessing the critical components of the IoT network. This dynamic mechanism of access control was very important in mitigating the dangers of security risks while not adversely affecting the whole network in case some devices became compromised.

AWS IoT Core was in charge of controlling communications between devices. The requirement for reliable, lightweight communication that makes use of the MQTT protocol and introduces safe transmission of data.

Combining AWS IoT Core with AWS Lambda provides all of the required cloud infrastructure to support aggregation of model updates while maintaining latency as low as possible and providing for secure authentication [22][7]

V. EXPERIMENTAL RESULTS

The proposed system was tested in two environments:- a smart home scenario and an industrial setting each designed to emulate real-world IoT deployments. These environments present decidedly different types of data and challenges that could be utilized for the validation of the effectiveness and scalability of the federated learning model.

- 1) *Smart Home Experiment:* The IoT network in this smart home mainly comprised temperature and motion sensors installed in almost every room of the house. The system looked out for anomalies, such as a malfunctioning HVAC system if the readings were hotter or cooler than normal recorded temperatures. It picked movements at night time since no one was expected inside the house. The federated learning system continued to improve over time regarding anomaly detection with fewer false positives as a result of more and more devices contributing to their model updates. Initially, it included harmless temperature changes in its list of anomalies but started doing better in distinguishing real security threats from benign environmental changes as it received continuous updates coming from other devices. Model accuracy increased by 15% over a 30-day testing period.
- 2) *Industrial Experiment:* In the industrial application, temperature sensors attached to crucial machinery plus activity monitors of network traffic monitoring device communications are part of the IoT network. The system noticed an infrequent surge in network traffic from one compromised sensor that turned out to have been infected by malware sending false data into the network.

This trust scoring mechanism effectively isolated the compromised sensor, cutting off any further malicious activity from that sensor. The sensor's trust score dropped significantly as soon as the system flagged its suspicious behavior. Therefore, the federated learning process automatically excluded the compromised device from contributing to the global model, avoiding model poisoning. The experimental result showed perceivability of dynamic adjustment of trust scores by the always changing behavior of the devices, along with network integrity due to isolation of compromised devices. Low latency was maintained by the experiment during incorporation of many more devices to the network and thus well scaling of the system indicated high scalability and robustness.

VI. CONCLUSION

The research successfully developed a Federated LearningBased Authentication and Trust Scoring System for IoT networks, which addresses critical challenges in scalability, privacy, and security. By utilizing federated learning, the system preserves data privacy while decentralizing anomaly detection. The integration of a dynamic trust scoring mechanism further strengthens the system's ability to mitigate security threats and manage device behavior effectively.

Through the smart home and industrial experiments, the system demonstrated its efficacy in detecting anomalies and dynamically adjusting trust scores to isolate compromised devices. Future work will focus on optimizing the federated learning process for larger, more complex IoT networks and exploring hybrid models that combine federated learning with other decentralized security approaches.

REFERENCES

- [1] Ng, W. W. Y., & Wong, K. W. (2017). A review of IoT security: Challenges and solutions. *IEEE Internet of Things Journal*, 4(5), 1082-1093.
- [2] Mukherjee, M., Matam, R., Shu, L., Maglaras, L., Ferrag, M. A., Choudhury, N., & Kumar, V. (2017). Security and privacy in fog computing: Challenges. *IEEE Access*, 5, 19293-19304.
- [3] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 1273-1282.
- [4] Kairouz, P., McMahan, H. B., Avent, B., et al. (2021). Advances and Open Problems in Federated Learning. *Foundations and Trends® in Machine Learning*, 14(1-2), 1-210.
- [5] Ryu, M., & Kim, Y. (2018). A trust scoring framework for IoT devices in edge computing environments. *International Journal of Advanced Computer Science and Applications*, 9(8).
- [6] Liu, H., Kumar, P., Li, T., & Wendt, J. (2020). Trustbased IoT network security: A review of trust management protocols and schemes. *IEEE Access*, 8, 207371-207394.
- [7] Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated learning with non-IID data. *arXiv preprint arXiv:1806.00582*.
- [8] Hsu, T., Qi, H., & Brown, M. (2019). Federated learning for mobile edge computing: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(3), 20312063.



- [9] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2), 1-19.
- [10] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637-646.
- [11] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645-1660.
- [12] Roman, R., Najera, P., & Lopez, J. (2011). Securing the Internet of Things. *Computer*, 44(9), 51-58.
- [13] Cao, Y., Chen, C., & Wang, Z. (2018). Anomaly detection in IoT systems: A federated learning approach. *IEEE Transactions on Mobile Computing*, 17(4), 911-925.
- [14] Alrawais, A., Alhothaily, A., Hu, C., & Cheng, X. (2017). Fog computing for the Internet of Things: Security and privacy issues. *IEEE Internet Computing*, 21(2), 34-42.
- [15] Hard, A., Rao, K., Mathews, R., Ramaswamy, S., Beaufays, F., Augenstein, S., ... & Simcha, D. (2018). Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604*
- [16] Cho, J. H., Swami, A., & Chen, R. (2011). A survey on trust management for mobile ad hoc networks. *IEEE Communications Surveys & Tutorials*, 13(4), 562-583.
- [17] Chen, R., Bao, F., Chang, M., & Cho, J. H. (2012). Trust management for secure cognitive radio ad hoc networks. *IEEE Network*, 26(1), 18-24.
- [18] Zhou, Y., Qin, Z., & Wu, Q. (2019). A blockchain-based federated learning model for secure and privacy-preserving mobile edge computing. *IEEE Access*, 7, 42559-42570.
- [19] TensorFlow Federated. (2020). TensorFlow Federated: Open-Source Framework for Federated Learning. Available at: <https://www.tensorflow.org/federated>
- [20] PySyft. (2020). A Library for Encrypted, Privacy-Preserving Machine Learning. Available at: <https://github.com/OpenMined/PySyft>
- [21] Amazon Web Services. (2020). AWS Lambda: Run Code without Thinking about Servers. Available at: <https://aws.amazon.com/lambda/>
- [22] Amazon Web Services. (2020). AWS IoT Core: Connect Devices to the Cloud. Available at: <https://aws.amazon.com/iot-core/>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)