Federated Learning on 5G Edge for Industrial Internet of Things

Xiaoli Liu, Xiang Su, Guillermo del Campo, Jacky Cao, Boyu Fan, Edgar Saavedra, Asunción Santamaría, Juha Röning, Pan Hui, and Sasu Tarkoma

Abstract

Industry 4.0, leveraging the Internet of Things (IoT) and Artificial Intelligence (AI), is a key enabler for many automated processes in modernized industrial applications. This paper addresses significant challenges pertaining to sensing and data analytics by connecting a large number of industrial IoT (IIoT) devices and deploying federated learning on 5G edge networks. We envision a federated learning-based 5G edge architecture for IIoT and develop an AI algorithm, i.e., an LSTM autoencoder algorithm for anomaly detection, on the 5G edge. We conduct comprehensive scalability analytics of communication and computation resources on our 5G edge IoT testbed. Our experimentation verifies that 1) federated AI algorithms can be deployed on 5G edge servers for latency-sensitive analytics, and 2) 5G edge supports scalable deployment of IIoT devices with low latency.

Introduction

Industry's transition to the digital world has already reaped several benefits, such as fully automated processes and predictive maintenance, which have paved the way for novel service developments and business models. Industry 4.0, supported by the Internet of Things (IoT) and machine intelligence, is a key enabler for many automated processes in modernized industrial applications. IoT enables production monitoring and controlling by collecting data from numerous sensors to increase manufacturing productivity, logistics, and other industrial contexts. Machine intelligence applies Artificial Intelligence (AI) algorithms to understand processes in industrial plants, predict events, and eventually support decision-making.

5G with low latency and high bandwidth plays a crucial role in delivering the data required by AI algorithms for real-time analytics. Additionally, 5G offers the capacity to connect many devices that intermittently transmit data. Edge servers, typically co-located with 5G base stations, allow for AI algorithms to be developed and deployed in proximity to sensors, actuators, and end-users [1]. By processing data collected from resource-constrained IoT devices leveraging computation-intensive AI algorithms at edge, Industrial IoT (IIoT) systems enables efficient decision-making by reducing contact frequency with cloud servers, thus reducing roundtrip delay; adhering to local identity management and access control policies; reducing lower communication costs through local processing, and load balancing between the application and network requests based on changes in the edge or core infrastructure, as well as adapting to temporary failures or maintenance.

The proliferation of sensors and connected devices in IIoT has brought heightened privacy concerns, as it necessitates the secure handling of vast amounts of sensitive operational data. For example, IoT-enabled medical devices collect real-time data on users' vital signs and health parameters, which are essential for tracking users' health conditions. However, due to the potential privacy risks, users may be hesitant to allow ML algorithms on centralized servers to analyze their personal data. Therefore, it would be ideal for storing sensitive data on users' trusted edge servers instead of a remote or cloud server. Federated learning (FL), a distributed ML approach, enables the training of ML models on trusted edge servers without transferring data to centralized cloud servers, which enhances the data owner's privacy [2]. Meanwhile, FL is well suited for edge computing applications and can leverage the computation power of edge servers [3]. Significant fundamental challenges exist in designing and implementing distributed FL architectures and systems [4] for IIoT while fully exploiting the computation and communication capacities of the 5G edge. Lu et al. [5] Have proposed integrating blockchain into FL to enhance security and privacy, where blockchain offers permission control for user participation and data encryption, in application scenarios beyond 5G. Luo et al. [6] Target optimization approaches of heterogeneity challenges of FL in 6G network by designing incentive mechanisms, network resource management, and personalized FL approaches. Limited research efforts have been made on deploying FL on 5G edge nodes for IIoT.

Our contributions are twofold. First, we propose an FL-based 5G edge architecture for IIoT connecting a large number of industrial

Xiaoli Liu (corresponding author), Jacky Cao, Boyu Fan, and Sasu Tarkoma are with the Department of Computer Science, University of Helsinki, 00014 Helsinki, Finland; Xiang Su is with the Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway; Guillermo del Campo, Edgar Saavedra, and Asunción Santamaría are with CEDINT, Universidad Politécnica de Madrid, 28040 Madrid, Spain; Juha Röning is with the Biomimetics and Intelligent Systems Group, University of Oulu, 90014 Oulu, Finland; Pan Hui is with the Department of Computer Science, University of Helsinki, 00014 Helsinki, Finland, and also with the Computational Media and Arts Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511442, China. Xiaoli Liu and Xiang Su are co-first authors.

Digital Object Identifier: 10.1109/MNET.2024.3469988
Date of Current Version: 14 January 2025
Date of Publication: 3 October 2024

Industry 4.0, supported by the Internet of Things (IoT) and machine intelligence, is a key enabler for many automated processes in modernized industrial applications.

resource-constrained IoT devices and develop FL with an LSTM autoencoder algorithm for anomaly detection on the 5G edge. To the best of our knowledge, this is one of the first efforts to enable FL on the 5G edge for IIoT. Second, we further demonstrate the feasibility of deploying FL with AI algorithms on a real-world 5G Test Network (5GTN) and conduct comprehensive scalability analytics of communication latency, throughput, and computation resources on 5G edge servers.

FL on 5G Edge for IIoT: Challenges and Enabling Technologies

Challenges

We have witnessed the proliferation of IoT devices in Industry 4.0. Numerous devices in industrial processes monitor and control production by creating a digital twin of the product being realized or the operation being achieved to increase manufacturing productivity, logistics, and many other industrial contexts. The core of the distributed automation systems in Industry 4.0 is reliable information exchange and decision-making [7]. In greater detail, we consider the following critical challenges:

- **Ubiquitous, High-Speed, and Reliable Connectivity for IIoT Devices.** IIoT networks present unique challenges compared to conventional IoT networks due to the critical nature of industrial operations and the complexity of industrial environments. Firstly, IIoT applications often require real-time monitoring and control, demanding low latency and high reliability to support critical processes. Secondly, industrial environments can be noisy, with various sources of electromagnetic interference that can degrade wireless communication signals. IIoT networks must overcome signal attenuation and interference challenges to maintain reliable communication between devices. Besides, IIoT networks have to scale effectively to accommodate the growing number of connected devices without sacrificing performance. Novel high-speed and reliable communication technologies are required to provide connectivity to heterogeneous, multi-vendor devices, enable interoperability by offering common software interfaces and compatible protocols, and handle data heterogeneity.
- **Big Data Streams Processing Capacities.** IIoT data is a type of big data, which is both large in scale and volume and is also continuous, often with rich time and location dependencies [8]. The potential subsequent integration of multiple sources further amplifies this challenge. Analyzing big data with AI algorithms extracts higher levels of information, guides the understanding of complex situations, and enables real-time analytics to provide user insights. Furthermore, data storage, management, confidentiality, and security also introduce significant challenges.
- **Distributed Latency-Sensitive, Privacy-preserving, and Reliable Training and Inference.** IIoT devices often have constrained and heterogeneous resources, complicating the deployment of complex ML tasks. Privacy-preserving training and inference ensure the protection of user data, but further amplify this challenge. Latency-sensitive decision-making generates insights timely before becoming obsolete. Reliable decision-making typically enforces the system to process significant amounts of data with interpretable AI models. Therefore, efficiency, privacy, and reliability are crucial, and different data models and data aggregation strategies in distributed training and inference must be considered.

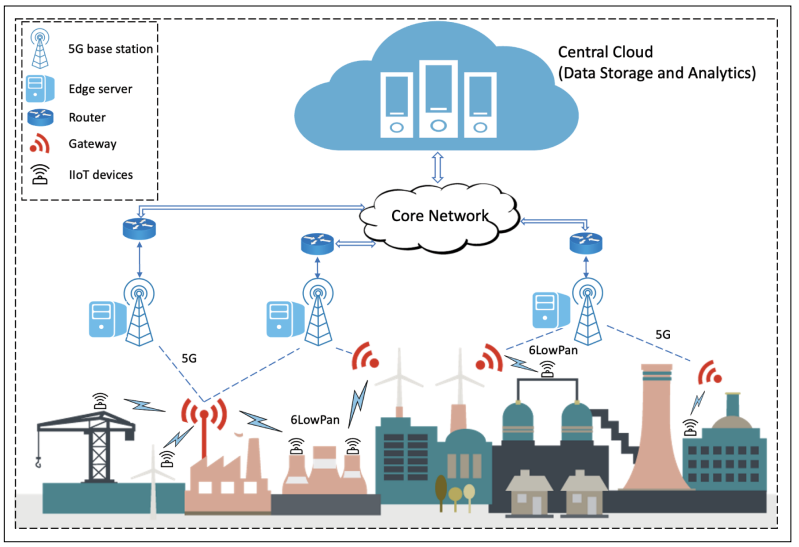

FIGURE 1. A conceptual FL 5G edge intelligence architecture for IIoT. Large volumes of data generated by IIoT devices in industrial processes is delivered to edge servers at their proximity for fast data analytics with reliable and high-speed data connections through gateways. 5G and 6LowPan networks allow for massive numbers of connections and provide support for large data transmission. Intelligent FL algorithms facilitate learning and decision-making at the edges of 5G networks by keeping measured data on edges without sharing with the cloud server. The ML model parameters are sent to cloud server for model aggregation.

Enabling Technologies

AI-powered systems are envisaged to overcome the emerging challenges of Industry 4.0 by fully unleashing the potential of edge intelligence. Figure 1 presents a conceptual architecture of FL on the 5G edge for IIoT.

Reliable IIoT Networks. Reliable decision-making depends on reliable data from reliable IoT networks. The reliability of IIoT networks requires minimum losses and a low, bounded latency. Table 1 presents the experimental results from the main IIoT network technologies: Zigbee, NB-IoT, 6LoWPAN, LoRaWAN, Sigfox, BLE and WiFi. Results show the average of 1000 messages that have been obtained by means of a validated Testbed device [9]. The Stability function determines the reliability of a communication technology, accounting for the variability of the latency and the number of losses. Energy consumption values given in milliwatts-second (mWs) represent the specific consumption of only the message

transmission for an IoT node. The other energy consumptions are not relevant to compare since they are all the same: a microcontroller in sleep or duty mode with the wireless transceiver deactivated. Note that Sigfox consumption is orders of magnitude higher due to its slow transmission rate (10 seconds on air for a single message) and distance to the receiving end. In fact, energy consumption increases as the range of coverage gets wider although not linearly due to modulation scheme, message size, frequency band, etc. Among them, IPv6 over Low-Power Wireless Personal Area Networks (6LoWPAN) provides a balance between technical properties (latency, error rate) and implementation features (energy consumption, cost). 6LoWPAN enables the use of IPv6 over the IEEE 802.15.4 standard, allowing each IoT device to be accessible via IP address [10]. The IEEE 802.15.4-2015 revision of the standard defines three MAC protocols, including Low Latency Deterministic Network (LLDN), Deterministic and Synchronous Multi-channel Extension (DSME), and Time Slotted Channel Hopping (TSCH), all of which are suitable for industrial applications [11]. LLDN is intended for single-hop low latency networks, such as factory automation. DSME is designed for highly reliable mesh networks, such as predictive maintenance. TSCH provides multi-hop and multi-channel communications, such as process control. Alternatives at the routing layer are mesh protocols, providing higher range and reliability, and star configuration, providing lower latency.

5G Networks. 5G is an M2M type communications enabler, featuring up to 10 Gbps speed, 1 ms latency, and 100% coverage and reliability. 5G networks aim to support IIoT and provide quality of experience-aware services for industrial applications. The dense deployment of 5G base stations, equipped with antenna arrays, significantly increases the available line-of-sights between IoT devices and 5G antennas. 5G networks have three characteristics, including massive machine-type communications (mMTC), enhanced mobile broadband (eMBB), and ultra-reliable and low-latency communications (uRLLC). 5G networks, with their mMTC characteristic, fulfill the requirement of ultra-dense machine communications by supporting connection densities of one million devices per square kilometer and fulfilling certain quality of service requirements. Thus, mMTC enables the connection of thousands of IoT devices simultaneously. eMBB supports high data rates (exceeding 10 Gbps), which fulfills the throughput requirement when thousands of IoT devices transmit large volumes of data simultaneously and may require gigabytes of network capacity. Finally, 5G offers uRLLC, which fulfills the requirements for extremely low latency communication, allowing for fast data transmission and control messages. Moreover, 5G networks are designed to enhance energy efficiency through lower power consumption for communication compared to other technologies, e.g., LTE-4G networks. Hence, equipping IoT devices with 5G will drastically decrease power requirements.

FL on Edge. FL is a distributed ML paradigm where a consortium of devices contributes collaboratively to a global neural network model instead of centralizing the data to train a global model. Edge intelligence [12] for IIoT requires integrating heterogeneous IoT data streams, analyzing IoT data with AI algorithms on edge devices, and deriving systemlevel understanding and knowledge for decision-making. FL on edge servers is optimal for IIoT due to its ability to process data locally on edge devices, diminishing the necessity for data transmission to a central server. This minimizes latency and enables real-time decision-making. Additionally, FL guarantees data privacy and security by keeping sensitive information on-premises, addressing compliance concerns. Its distributed nature also enhances scalability and system robustness, as model training can continue even if certain devices are offline. FL on 5G edge networks harness the advantage of 5G such as high bandwidth, low latency, and increased connectivity, to improve the efficiency and effectiveness of FL system for improved participation, fast model updates and inference, continuous model training without disruption, and enhanced model robustness, which further enable reliable IIoT applications.

FL guarantees data privacy and security by keeping sensitive information on-premises, addressing compliance concerns.

Technology	Zigbee	NB-IoT	6LoWPAN	LoRaWAN	Sigfox	BLE	WiFi
End-to-end latency (ms)	48	1797	22	397	3695	27	32
Latency standard deviation (ms)	5	1352	9	5	294	13	9
Error Rate (%)	0	0	0.02	0.66	0	0	0
Stability (0–1)	0.592	0.036	0.924	0.993	0.922	0.002	0.036
Energy consumption per message transmission (mWs)	12.57	44.40	0.92	31.65	4024	5.05	17.02

TABLE 1. Comparison of IoT network technologies.

Edge Intelligence Algorithm: A Case Study of Anomaly Detection

FL Architecture for Anomaly Detection

In this section, we present the development of FL algorithms, i.e., anomaly detection for IIoT data, on edge servers in proximity to IoT devices. Anomaly detection plays an important role in IIoT, such as facilitating early identification of deviation from normal operations in industry machinery, supporting predictive maintenance strategies, and identifying unusual network behaviors within IIoT systems. Implementing federated anomaly detection in edge devices for IIoT data offers several significant benefits. Firstly, it allows for real-time analysis without the necessity of transmitting raw data to a centralized server, thereby reducing network bandwidth usage and latency, enabling quicker anomaly detection and response. Secondly, it enhances data privacy and security by keeping sensitive information localized to the edge devices, mitigating the risk of data breaches and ensuring compliance with privacy regulations. Furthermore, federated anomaly detection

improves scalability and fault tolerance, leveraging processing power of edge devices.

Figure 2 presents an overview of the federated anomaly detection architecture for energy monitoring. The collected sensor readings on different edges are denoted by D_1, D_2, …, D_k, where k is the number of edge nodes. The amount of computation is controlled by: C, the fraction of edge nodes that perform computation on each round ($C = 1$ corresponding to full-batch gradient descent meaning all the edge nodes participating in the training process); E, the number of training passes each edge node makes over its local dataset on each round; and B, the local minibatch size used for the edge node updates ($B = \infty$ indicates that the fully local dataset is treated as a single minibatch).

Model

Our anomaly detection model is based on reconstructing the sensing data shown in the top left corner in Figure 2. We design an LSTM autoencoder architecture, a stacked LSTM network consisting of LSTM layers and TimeDistributed dense layers, for energy consumption anomaly detection. Our collected data from IIoT devices is temporal with continuous time series exhibiting periodic (or cyclic behavior) patterns and showing long-term trends. Anomaly is often defined as long-term trends and it is essential that the anomaly detection model could capture the dependence and patterns over continuous time steps. LSTM is selected due to its ability to learn long-term dependence and handle inputs of varying lengths which make it particularly suitable for sequence modeling problems. Meanwhile, the features of LSTM make it easy to train under FL architecture compared to alternative anomaly detection methods. Considering a time series $X = \{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\}$, where each point $x^{(t)}$ in the time series refers to an observed value at time t. The model expects inputs of a sequence with K time steps and outputs a sequence with K time steps. The reconstruction errors between reconstruction values X' and real measurements X from sensors are minimized. Specifically, the mean absolute error (MAE) is defined as the loss function for finding the optimized parameters. Adam optimization algorithm with hyperparameters setting $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ is deployed for the LSTM autoencoder. Adam is an adaptive learning rate optimization algorithm that combines the advantages of AdaGrad and RMSProp and has many appealing qualities such as computational efficiency and invariant to diagonal rescale of the gradients. It is widely regarded as being fairly robust to the choice of hyperparameters [13]. On each edge node, the data is partitioned into two parts: the former 70% of the whole data set is used for training, and the remaining 30% is used for testing. Meanwhile, 10% of the training set is used for validation when the best parameters are selected. The reconstruction errors on the training data set are used to formulate the threshold for anomaly detection. A data point on a data set is labeled as an anomaly if the reconstruction error for the data point is greater than the threshold.

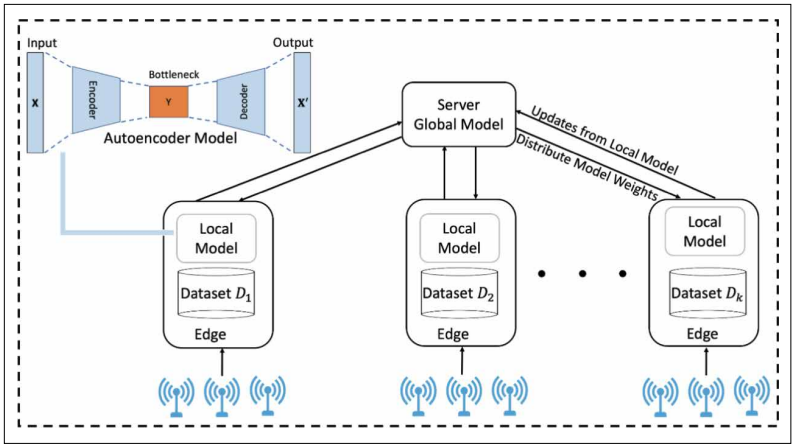

FIGURE 2. Federated architecture for anomaly detection where federating participating devices using local data to train a global anomaly detection model instead of sending the data to a central server. The cloud server distributes the current model parameters to selected edge nodes in each training round. Each selected edge node locally trains and updates the local model's parameters by calculating the stochastic gradient descent (SGD). Then the server takes a weighted average of the updates for global model's updates.

Experiments and Results Analysis

We use data from the CeDInt IoT network [14] to design and evaluate the performance of our proposed federated algorithm for anomaly detection. The CeDInt IoT network monitors energy consumption and ambient parameters (temperature, humidity, and presence) within the CeDInt building, including HVAC, lighting and other systems. The federated anomaly detection model is used to detect anomalies within HVAC systems by analyzing related energy consumption and ambient data.

We consider three model structures: 1) two layers with 32 units; 2) two layers with 64 units; and 3) four layers with 64, 32, 32, 64 units. Meanwhile, in each model structure, we fix $C = 1$, $E = 1$, and add more computation per client on each round by decreasing B [2]. Figure 3(a)-(c) present the loss curves of the three model structures with different combinations of (B, E). Our results demonstrate that 1) adding more local SGD updates per round can significantly decrease training loss. The expected number of updates per client per round is $u = En/B$. We see that increasing u by varying B is effective. We also note that decreasing B is taking advantage of available computation resources of edge devices and this in practice, should be a primary parameter to be tuned; and 2) the model with two LTSM layers with both 64 units has the best learning effect, which means that the neural network structure also affects the training results.

We demonstrate how to use a federated model to identify the anomaly, specifically formulating a threshold. As we assume that all the data points on the training data set are normal, and the max value of the reconstruction errors on the training data set is the worst for our proposed model to reconstruct the data point. Therefore, it is suitable to be used as the threshold. Figure 3(d)–(f) present the process of detecting anomalies in the HVAC scenario. The MAE distribution on the scaled training data set is shown in Figure 3(d), and the max MAE value of 0.3 is used for the threshold. Figure 3(e) presents the dataset where

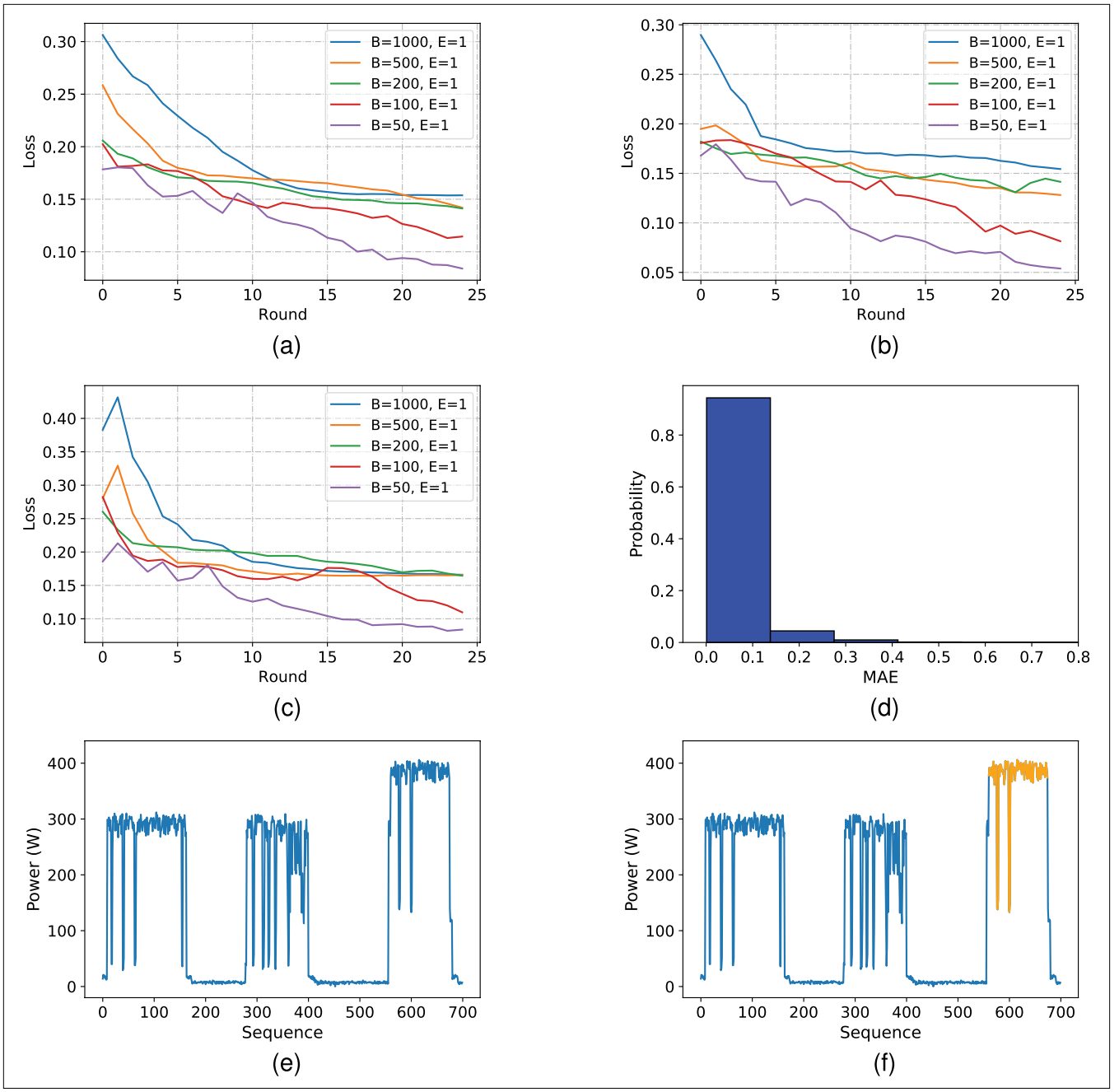

FIGURE 3. a)–c) Performance of the proposed federated anomaly detection model with different model structures with ($B = 1000$, $E = 1$), ($B = 500$, $E = 1$), ($B = 200$, $E = 1$), ($B = 100$, $E = 1$), ($B = 50$, $E = 1$); d) MAE on the scaled training data set in HVAC scenario; e) an example of the testing dataset containing anomaly; and f) anomaly detected marked with orange color using the proposed model on testing dataset e).

a sudden jump is added to the original dataset, which is used to evaluate the performance of the proposed model and show whether the proposed model could detect the sudden jump as an anomaly and to what extent. The difference is that the values for the time interval [560, 680] in Figure 3(e) suffer from a sudden jump with 100 W more when compared to the normal testing data set. We calculate the true positive (TP), false negative (FN), false positive (FP), and true negative (TN), and use four performance measures, namely precision $\left(\frac{TP}{TP+FP}\right)$, recall $\left(\frac{TP}{TP+FN}\right)$, F1 score $\left(\frac{2TP}{2TP+FP+FN}\right)$, and detection accuracy $\left(\frac{TP+TN}{TP+FN+FP+TN}\right)$ to evaluate the proposed federated model. *TP* is the number of points that lie in time interval [560, 680] detected as anomalies, and FN is the number of points that lie in time interval [560, 680] detected as normal points. Similarly, FP and TN are the numbers of points in time interval [0, 559] detected as anomalies and normal points, respectively. The precision, recall, F1 score, and detection accuracy of the proposed federated detection model on the dataset shown in Figure 3(e) are 0.954, 0.983, 0.968 and 0.950, respectively, which demonstrate that the proposed federated model can effectively detect anomalies. Figure 3(f) shows the detected anomalies in time interval [560, 680] by using the proposed

federated model marked with an orange color, which further demonstrates the effectiveness of the proposed federated model.

5G Edge Experimentation and Analytics

Testbed

This section presents the performance of deploying FL model on a real-world 5G edge IoT testbed. Our testbed is composed of IoT networks and 5G edge testbed (5GTN).

IoT Network: The IoT network at the CeDInt building is used 1) to evaluate 6LoWPAN reliability and compliance with industrial constraints and 2) to collect IoT data for anomaly detection algorithm deployable on edge servers. Based on previous experimental evaluations, 6LoWPAN outperforms alternative protocols regarding communication latency: 6LoWPAN-measured single node latency is 20 ms, which outperforms BLE (˜26 ms), WiFi (˜32 ms), Zigbee (˜40 ms), LoraWAN (˜290 ms), and Sigfox (˜3.7 s). Besides, the mesh network topology allows for a dynamic routing configuration, increasing communication range while reducing overhead and error rate. Figure 4(a) presents different elements of the experimental IoT testbed.

5GTN: 5GTN is a full-scale 5G micro operator, providing both standalone and non-standalone 5G and LTE connectivity. We conduct our experimentation at the University of Oulu implementation of 5GTN, which has air interfaces of two 5G macrocells (n78), several LTE macrocells (B28, B7, B42), and a LoRa network supporting frequencies [0.7, 2.1, 2.6 and 3.5] GHz. Moreover, our testbed supports heterogeneous wireless technologies, including IEEE 802.11, Bluetooth LE, LoRa, NB-IoT, UWB and LTE evolutions. Edge servers are deployed on 5GTN to support latency-sensitive data analytics. Figure 4(b) presents the 5GTN cell tower and the Multi-access Edge Computing (MEC) server. The edge server in this experimentation has an Intel Core i7-8700 CPU, 32 GB memory, and an NVIDIA GeForce RTX 2080 Ti GPU. To support these experiments, we deploy a Stockholm-based cloud Amazon EC2 server, which has comparable specifications as our edge server, i.e., 8 virtual CPUs, 32 GB memory, and an NVIDIA T4 GPU.

A PC (Intel Core i7-8700K CPU, 32 GB memory, and an NVIDIA GeForce RTX 2080 with Max-Q Design GPU) relays the real collected data from IoT devices. With this approach, we can evaluate massive deployments of devices through multiple parallel threads (one thread representing one device). Data is sent from each thread to a MEC server connected to 5GTN. The PC relaying the data can access the network through a 5G modem.

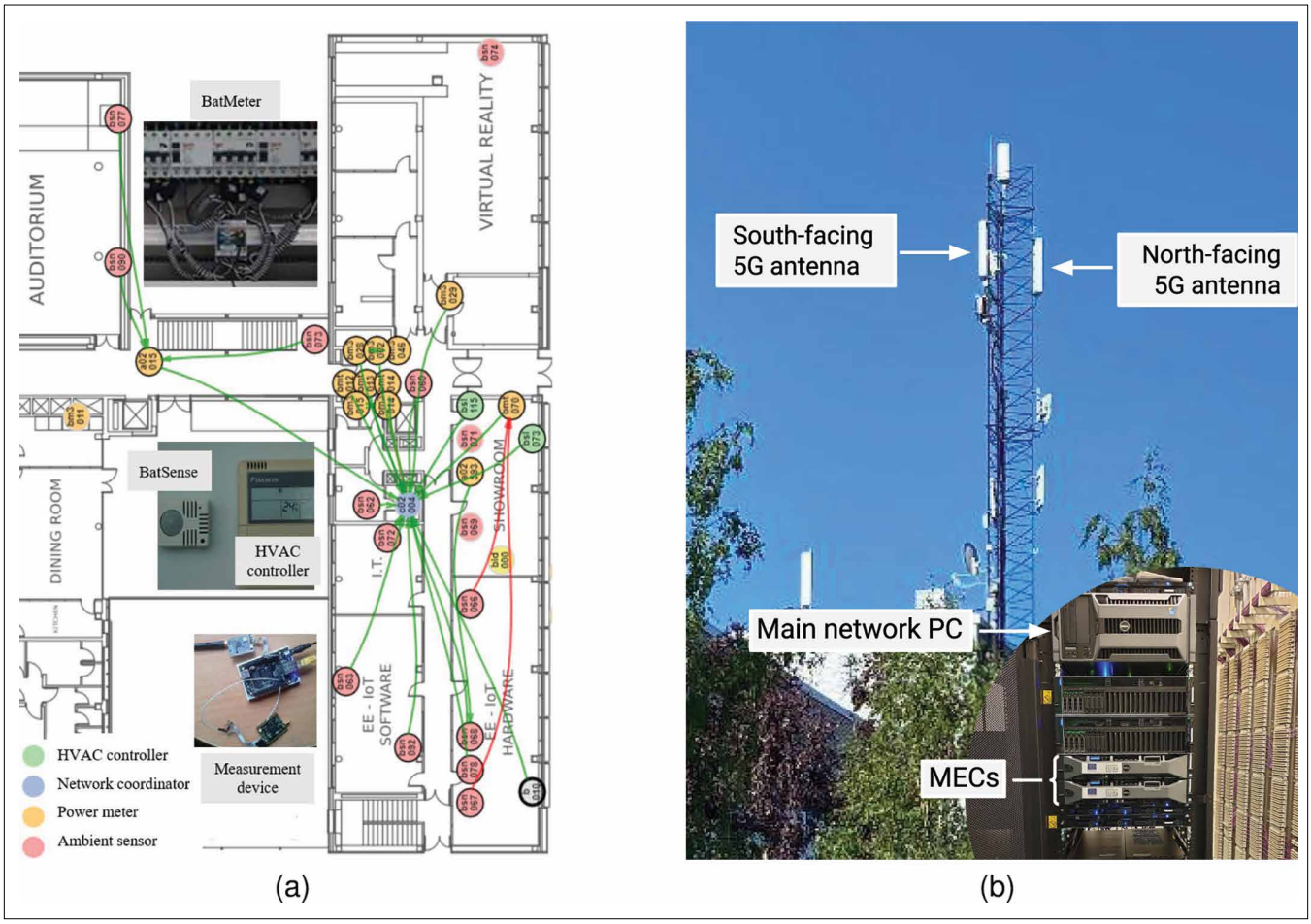

FIGURE 4. a) Floor map of IoT sensor deployment in CeDInt building, with the location of the different IoT nodes and communication paths, power meters (BatMeter) installed at electric panelboards, ambient sensors (BatSense) close to a HVAC controller, and the measurement device. b) 5GTN with MEC server in our experiments.

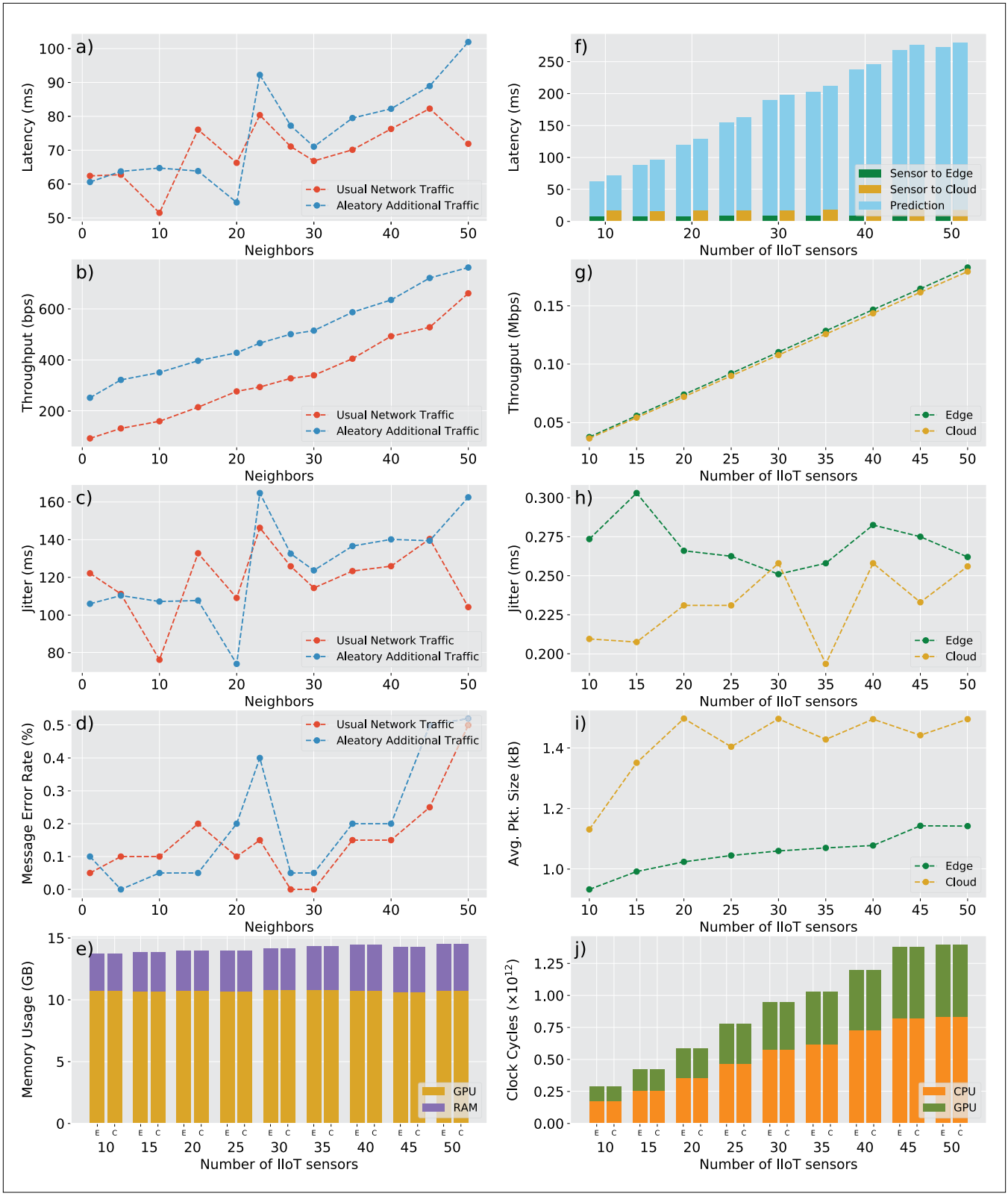

FIGURE 5. Left: 6LoWPAN results of the a) latency, b) throughput, c) jitter, and d) error rate. The red lines represent the usual network traffic (configuration messages, periodic measurement requests from higher-level applications, and event-triggered message transmissions). We add additional aleatory traffic (57 byte messages every 4 seconds, blue lines) to simulate real-world environments and test reliability [15]. Then e) are the results of the GPU memory and RAM usage for the edge- (E) and cloud- (C) deployed algorithm, with the cloud server requiring an average of 12.7% more memory resources than the edge. Right: f) cellular 5G latency and the computation time for prediction where the data connection to the edge server has an average latency of 10.5 ± 2.0 ms, while the connection to the cloud is approximately 40.3% larger at 17.6 ± 0.8 ms, and the prediction latency for both edge and cloud servers is similar, with an average percentage difference of 1.30%. Then the cellular 5G results of the g) throughput, h) jitter, and i) average packet size are presented; as well as j) the GPU and CPU clock cycles results for the edge- (E) and cloud- (C) deployed algorithm. The edge executes 9.62% more complete clock cycles than the cloud server.

5G edge intelligence combines advanced connectivity, compact processing power, and AI at the edges of 5G networks, which presents an emerging trend for Industry 4.0.

Results and Analytics

To analyze the performance of deploying the FL model on the 5G edge for IIoT, we study:

- Data transmission latency, throughput, error rate, and jitter. We separate overall latency into the latency of transmitting data from IoT sensors to a 5G modem with 6LoWPAN and the latency of forwarding the data from the 5G modem to an edge server with a 5G connection.
- Scalability, by increasing the number of emulated devices to deliver data from a 5G modem to an edge server.
- Required computation resources (CPU, GPU, and memory) by the FL trained LSTM autoencoder algorithm on edge servers.

We analyze different communication parameters while increasing the number of neighbors from 1 to 50 in increments of 5. Figure 5(a)–(d) presents results of the 6LoWPAN experimentation. Node hopping increases average latency values compared to single communications (60-70 ms vs 20 ms). Latency and jitter value oscillations are caused by nodehopping and aleatory event-triggered messages and added traffic. We observe that the message error rate is less relevant (below 0.5%). Experimental results of latency, jitter, and error rate validate that utilizing 6LoWPAN in industrial applications ensures high-speed, reliable IIoT communications. For experimentation of FL on the 5G edge, we measure the latency, throughput, jitter, and average packet size from sending data from the sensors to a receiving MEC server and a cloud server with a 5G network connection. We vary the number of sensors from 10 to 50 in increments of 5. Figure 5(f)–(i) present the results of the 5G experimental tests. We observe that as the number of IoT sensors increases, the data transfer latency, throughput, and average packet size increase for both edge and cloud connections. For timely data transfer, an edge server is, therefore, the most suitable. For both connections to the edge and cloud server, the throughput has approximately the same values for each sensor number. The jitter shows more variability, e.g., increased jitter at 15 sensors at the edge and at 35 sensors for the cloud, which reflects the dynamic nature of a live cellular network. To test the capabilities of the proposed LSTM autoencoder algorithm, we deploy the algorithm on both an edge and cloud server and compare the required computation resources, i.e., the total prediction latency, memory usage, and the number of clock cycles during algorithm execution. Figure 5(f), (e), and (j) present the results of this experimentation.

Discussion and Open Challenges

5G edge intelligence combines advanced connectivity, compact processing power, and AI at the edges of 5G networks, which presents an emerging trend for Industry 4.0. Federated optimization enables increasingly complex networked systems involving heterogeneous devices, service providers, and network operators to become more intelligent and autonomous in network management with privacy protection. In this paper, we focus on sensing and data analytics by connecting many resource-constrained IIoT devices and deploying a federated AI algorithm on the 5G edge. Our contributions are twofold: 1) we envision FL on 5G edge-enabled IIoT architecture; and 2) we develop and deploy federated LSTM autoencoder anomaly detection on the 5G edge and conduct comprehensive scalability analytics of communication and computation resources on our 5G edge IoT testbed. Our experimentation verifies that 5G edge supports scalable deployment of IIoT devices with low latency and federated algorithm deployed on 5G edge servers for privacy-sensitive analytics.

This study has several open challenges that need further investigation. First, how to efficiently process continuous data streams locally arrived at 5G edge servers, which is vital for real-time analytic and decision making. Second, how to optimize the FL algorithm while to minimize the energy consumption during the FL model training and communication process if energy consumption of edge servers or devices is a significant concern. Third, how to address potential attacks of FL systems on 5G edge, such as gradient leakage and model poisoning attacks, is still a significant concern. Even data is kept locally without sharing and only parameter updates (gradients) are sent during the training process, it is still possible for the adversaries to disclose valuable information or even reconstruct the raw data from the leaked parameter gradients. Adversaries can also attack the model aggregation process to alter the parameters of the global model to corrupt the model, causing the global model to behave undesirably and produce inaccurate predictions.

Acknowledgment

This work was supported in part by the Academy of Finland under Grant 345008 and Grant 326305 and in part by the Nordforsk Nordic University Cooperation on Edge Intelligence under Grant 168043.

References

[1] Y. Liu et al., "Toward edge intelligence: Multiaccess edge computing for 5G and Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 6722–6747, Aug. 2020.
[2] B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artif. Intell. Statist.*, 2017, pp. 1273–1282.
[3] R. Yu and P. Li, "Toward resource-efficient federated learning in mobile edge computing," *IEEE Netw.*, vol. 35, no. 1, pp. 148–155, Jan. 2021.
[4] P. Kairouz et al., "Advances and open problems in federated learning," *Found. Trends Mach. Learn.*, vol. 14, nos. 1–2, pp. 1–210, Jun. 2021.
[5] Y. Lu et al., "Blockchain and federated learning for 5G beyond," *IEEE Netw.*, vol. 35, no. 1, pp. 219–225, Feb. 2021.
[6] B. Luo et al., "Optimization design for federated learning in heterogeneous 6G networks," *IEEE Netw.*, vol. 37, no. 2, pp. 38–43, Mar. 2023.
[7] P. K. Malik et al., "Industrial Internet of Things and its applications in Industry 4.0: State of the art," *Comput. Commun.*, vol. 166, pp. 125–139, Jan. 2021.
[8] X. Liu et al., "Enhancing veracity of IoT generated big data in decision making," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2018, pp. 149–154.
[9] E. Saavedra et al., "A universal testbed for IoT wireless technologies: Abstracting latency, error rate and stability from the IoT protocol and hardware platform," *Sensors*, vol. 22, no. 11, p. 4159, May 2022.

[10] J. Tournier et al., "A survey of IoT protocols and their security issues through the lens of a generic IoT stack," *Internet Things*, vol. 16, Dec. 2021, Art. no. 100264.
[11] L. Alkama and L. Bouallouche-Medjkoune, "IEEE 802.15.4 historical revolution versions: A survey," *Computing*, vol. 103, no. 1, pp. 99–131, Jan. 2021.
[12] D. Xu et al., "Edge intelligence: Empowering intelligence to the edge of network," *Proc. IEEE*, vol. 109, no. 11, pp. 1778–1837, Nov. 2021.
[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, Y. Bengio and Y. LeCun, Eds., May 2015, pp. 1–11.
[14] G. del Campo et al., "BatNet: A 6LoWPAN-based sensors and actuators network," in *Ubiquitous Computing and Ambient Intelligence*. Cham, Switzerland: Springer, 2012, pp. 58–65.
[15] M. Herlich and C. Maier, "Measuring and monitoring reliability of wireless networks," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 76–81, Jan. 2021.

Biographies

Xiaoli Liu (xiaoli.liu@helsinki.fi) is currently a Research Coordinator with the Department of Computer Science, University of Helsinki, Finland. Her research interests include federated learning, edge intelligence, and augmented reality.

Xiang Su (Member, IEEE) (xiang.su@ntnu.no) is currently an Associate Professor with the Norwegian University of Science and Technology, Norway. He has extensive expertise in the Internet of Things, edge intelligence, and extended reality.

Guillermo del Campo (Member, IEEE) (guillermo.delcampo@upm.es) is currently an Assistant Professor with the Universidad Politécnica de Madrid, Madrid, Spain. He is also the Head of the IoT/EE Research Group, CEDINT, Universidad Politécnica de Madrid. His research interests include the Internet of Things, wireless communications, visible light communications, and smart environments.

Jacky Cao (jacky.cao@helsinki.fi) received the M.Phys. degree from Durham University in 2019. He is currently pursuing the Ph.D. degree with the University of Oulu, Finland. His research interests include mobile augmented reality, edge computing, and 5G networks.

Boyu Fan (boyu.fan@helsinki.fi) received the master's degree from Beihang University, Beijing, China, in 2020. He is currently pursuing the Ph.D. degree with the Department of Computer Science, University of Helsinki, Finland. His research interests include pervasive computing, the Internet of Things, and federated learning.

Edgar Saavedra (edgar.saavedra@upm.es) received the master's degree in telecommunications engineering from the Universidad Politécnica de Madrid, Madrid, Spain. He is currently pursuing the Ph.D. degree with CEDINT, Universidad Politécnica de Madrida, with a focus on sensor wireless networks for energy efficiency in industry, homes, and cities.

Asunción Santamaría (asun.santamaria@upm.es) is currently a Professor with the Telecommunications School, Universidad Politécnica de Madrid, Madrid, Spain. She is also the Director of CEDINT, Universidad Politécnica de Madrida. She has extensive expertise in the Internet of Things, data communications, network communication, network architecture, cloud computing, and virtual and augmented reality.

Juha Röning (juha.roning@oulu.fi) is currently a Professor of embedded system with the University of Oulu, Finland. He is also a Principal Investigator with the Biomimetics and Intelligent Systems Group (BISG). His research interests include computer vision, robotics, intelligent signal analysis, and software security. He is currently serving as the Board of Director for euRobotics aisbl. He is also a Steering Board Member of ARTMIS-IA.

Pan Hui (Fellow, IEEE) (pan.hui@helsinki.fi) received the Ph.D. degree from the Computer Laboratory, University of Cambridge. He is currently a Chair Professor of Computational Media and Arts (CMA); a Chair Professor of Emerging Interdisciplinary Areas; and the Director of the Center for Metaverse and Computational Creativity and the HKUSTDT Systems and Media Laboratory (SyMLab), The Hong Kong University of Science and Technology (Guangzhou). He has extensive experience on augmented reality, virtual reality, and metaverse. He is an ACM Distinguished Scientist and a member of Academia Europaea.

Sasu Tarkoma (Senior Member, IEEE) (sasu.tarkoma@helsinki.fi) received the Ph.D. degree in computer science from the University of Helsinki, Finland, in 2006. He is currently the Dean of the Faculty of Science and a Full Professor with the Department of Computer Science, University of Helsinki. He is also the Director of the Helsinki Center for Data Science (HiDATA). His research interests include distributed systems, data analytics, mobile and ubiquitous computing, artificial intelligence, and 6G.