World Scientific Research Journal Volume 8 Issue 11, 2022
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

An Anomaly Detection Method Based on Federated Learning and Homomorphic Encryption on Cognitive Internet of Things for Fog-based Smart Home

Changjie Liu[1,*]

[1]School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

*Correspondence Author: lcj@ncepu.edu.cn

Abstract

With the deployment of more and more edge devices and the improvement of intelligent technologies level, research on the cognitive Internet of Things for fog-based smart home emerges as the times require. In this context, the study of abnormal network traffic also encountered new problems, such as data islands and user privacy leakage. In order to solve these problems, we propose an anomaly detection method based on federated learning and homomorphic encryption on cognitive Internet of Things for fog-based smart home. This model is a new data sharing mode, which only needs a kind of processing of shared data, not the shared data itself. Firstly, a model based on federated learning on cognitive Internet of Things for fog-based smart home is proposed to solve the problem of data island through multi model cooperation of different terminals. Secondly, to protect the user privacy, we use homomorphic encryption to achieve encrypted transmission. By using homomorphic encryption, data aggregation of model parameters is realized. After analysis, the model can be extended to multi-dimensional or multi-level. Finally, security analysis, model training performance, computational complexity and communication cost are investigated. The simulation results show that the proposed scheme behave well in both cost and performance.

Keywords

Fog-based smart home; Cognitive Internet of Things; Anomaly detection; Federated learning; Homomorphic encryption.

1. INTRODUCTION

Fog computing was first proposed by Bonomi et al. in 2012 [1], by adding fog nodes between cloud server and terminal devices to alleviate the pressure of computing, communication and storage in cloud computing paradigm. In this way, the computing efficiency, communication load, storage capacity and location-based service (LBS) capacity of the whole system are enormously improved. Fog computing is an extended application of edge computing, which has been a hot spot among scholars and industry attention [2]. In 2015, Yi et al. proposed a definitions of fog computing and a deeper comprehensive [3]. They focused on the fog computing platform designed, goals, analyzed the advantages and disadvantages with several exemplar applications. In 2014, Vaquero et al. understands the fog, cloud, peer-to-peer networks, sensor networks, network virtualization functions or configuration management techniques from different angles [4].

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

With the development of fog computing and Internet of things, fog-based smart home was proposed by Bansal et al. in 2017 [5]. In 2018, by using the concept of fog computing at the smart gateway, Verma and Sood proposed the remote patient health monitoring in smart homes [6]. The model proposed in this paper can realize intelligent service through application distributed storage, embedded data mining, and notification services at the sensor network. Rahimi et al. in 2020 proposed a novel Systematic Literature Review method for fog-based smart home [7]. In 2020, by using IoMT Technology, Bhatia focused on the remote surveillance of domestic animals' health conditions in the smart home environment [8]. All of the above solutions put forward their own ideas to solve the problems of smart home, but these solutions do not consider the cognitive characteristics of the Internet of things.

Anomaly detection is a typical application and research hotspot of smart home [9]. In 2013, Bhuyan et al. provided a structured which can become quickly familiar with every aspect of network anomaly detection [10]. In 2020, Nachman and Shih proposed a new unsupervised anomaly detection with Density Estimation (ANODE) technique by leveraging some breakthroughs in neural density estimation [11]. They constructed a fully data-driven likelihood ratio of data versus background, by estimating the conditional probability density data in a signal region. Fei et al. in 2020 thought the normal and the anomalous data are expected to be distinguishable based on restoration errors and wanted to break this equivalence by erasing selected attributes from the original data and reformulate it as a restoration task [12]. Garg et al. in 2020 proposed a multi-stage anomaly detection model by rectifying the problems incurred in traditional DBSCAN. These solutions do not consider the problem of data island and user privacy leakage, so they cannot solve the problem proposed in this paper.

In order to solve these problems, we propose an anomaly detection method based on federated learning and homomorphic encryption on cognitive Internet of Things for fog-based smart home. The main contributions of this paper are as follows.

1) Firstly, a model based on federated learning on cognitive Internet of Things for fog-based smart home is proposed to solve the problem of data island through multi model cooperation of different terminals.

2) Secondly, in order to protect the model parameters and other user privacy data, we use homomorphic encryption to achieve encrypted transmission. By using homomorphic encryption, data aggregation of model parameters is realized. After analysis, the model can be extended to multi-dimensional or multi-level.

3) Finally, security analysis, model training performance, computational complexity and communication cost are investigated. The simulation results show that the proposed scheme has good performance and low cost.

The rest of this paper is organized as follows. In section 2, some background knowledge is mentioned, such as cognitive Internet of Things and federated learning. In section 3, system model, security requirement and design goals are given. In section 4, the proposed scheme is introduced in detail. Security analysis and performance analysis are investigated in section 5 and 6, respectively. In section 7, related words is given. In section 8, we conclude our word of this paper.

2. PRELIMINARIES

Before introducing the proposed system model, some background knowledge needs to be mentioned, such as cognitive Internet of Things and federated learning.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

2.1. Cognitive Internet of Things

The rapid development of sensor technologies, cloud systems and emerging of 5G communication technologies makes IoT larger and more complex, which is well presented in the massive data in both varieties and amounts and busier and busier network traffic in every second. Such changes proposed new challenges on the IoT's ability of processing and decision-making and transmission, which promote the evolution of next generation IoT, Cognitive Internet of Things as well.

Cognitive Internet of Things was proposed by Wu et al. in 2014 [14] and has attracted the attention of many scholars [15-20]. By applying cognitive computing technologies and artificial intelligence to traditional IoT, Cognitive Internet of Things can provide better performance in summarizing the massive data produced by IoT devices and get estimations of current situation then make opportune decisions. More specially, cognitive computing, which enables computer systems make judgements and decisions in the way people do, results in saving time and effort, increasing calculating resource efficiency and higher accuracy.

The framework of Cognitive Internet of Things is shown in Figure 1, the bottom side of this framework composed of massive IoT sensors, processors and controllers, which is strictly exposed to physical environment, so that playing the role of raw data producers and terminals of IoT service. In this layer, large scale of data which may contain the details of someone's daily life routine or describe the sale situation of a shop is continuously produced. On one hand, the increasing scale of data as time goes by is a challenge to the calculating abilities of IoT devices, which calls for more powerful processors, controllers and storage, on the other hand, data from such devices may contains consumers' privacy such as one's electrical consumption trend over time so that further speculate his or her habits and professions, which should be specially protected in consideration of consumers' safety and privacy. The layer in middle plays the role of data transmission, which contains both wired and wireless methods and the protocols of it includes emerging protocols which are fit for communications of lightweight, movable and distribute devices such as CoAP, MQTT, DDS, and XMPP and traditional Internet protocols such as REST/HTTP. The progress of 5G enables IoT devices communicates in a higher data rate and mobility with lower latency and energy cost at the same time, which accelerating the density and sensitivity of IoT devices.

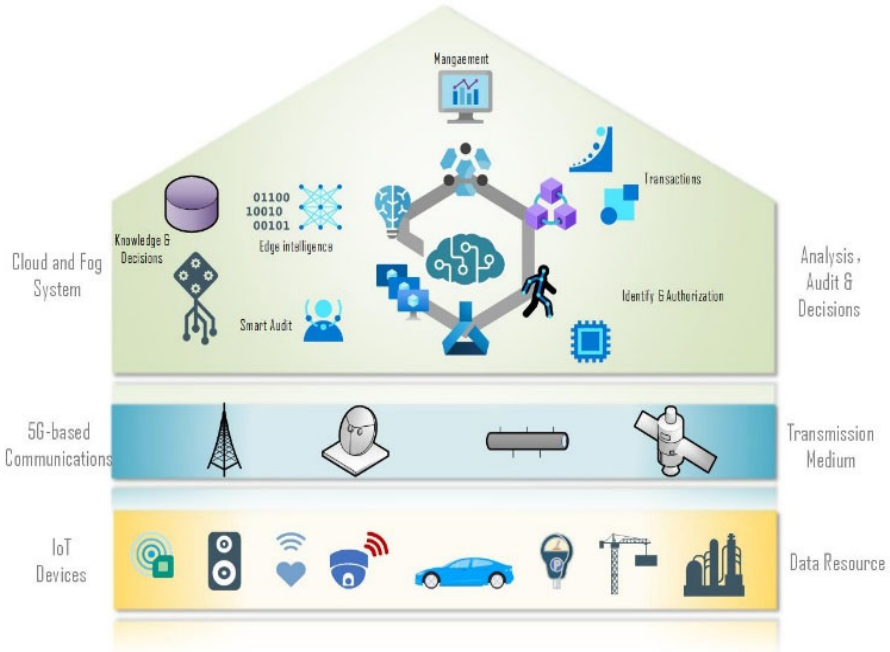

Figure 1. The framework of Cognitive Internet of Things

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

On the top of this framework lies the data center, which makes analysis and comparison of the data gain from bottom and then makes the corresponding decisions according to existing knowledge and the features of the event. With the development of cognitive computing, cloud system and edge intelligence, data center become efficient: By evaluating its need of calculating resource and timeliness and its imitator's location, the data center can and allocate tasks such as data aggregation and analysis and feature extraction to some fog nodes which lies between cloud and user terminals, in doing this ,on one hand the calculating resource of the whole system get reasonable distribution, avoiding the situation that tasks that need cloud's powerful computing abilities have to wait for the cloud finishes tasks that need less computing, on the other hand it can shorten the average response time of most tasks, as performing tasks on fog nodes will significantly save the time used for communication, which is an obvious improvement to tasks that require frequent communications and less computations.

The Cognitive Internet of Things brings more and more advantages, though, there are also many problems and deficiencies need to be addressed: (1) the lack of protection to consumers' privacy on IoT terminal devices (2) a widely applied standard in the context of 5G-based IoT transmission (3) better algorithm and center system structure that can be used to recognize varieties of user events and the way to organize and maintain the knowledge base.

2.2. Federated Learning

Advances in machine learning and edge computing prove to be an attractive option for IoT to analyses the behavior patterns of users so that it can provide better services, while, the isolated data island and privacy embedded in user data cannot be ignored. In consideration of different manufacturers of IoT devices, data in IoT have their own definition and structure, results in varieties of heterogeneous data that is hard to straightly put in use, also, trying to make an integration of all kinds of data seems to be an impossible mission. At the same time, privacy protection is getting more and more attention, major countries and organizations in the world have all issued relevant regulations such as EU's General Data Protection Regulation and China's Personal Information Protection Law, using consumers' data to train a machine learning model without permission may result in prosecution even if with the purpose to provide better service to them.

Under this circumstance, with the advantage of leveraging isolated data resources to train machine learning models without leakage of users' privacy, federated learning attracted a larger number of researchers' attention because of its subversive data sharing mode [21-26]. Federated learning was first proposed by Google in 2016 to solve the problem of mass terminal mobile phone system update [21], as shown in Figure 2, user data is only used for the training of local model, which reduces the risk of the leakage of privacy due to interception or man-in-the-middle attack in data transmission. After local training, local user keeps their own data safely and send upgraded model to global center, where each participants' model is aggregated and then sent back for further training. During the whole process, local data is never been transmitted and the transmission of local model also has methods to protected user's privacy such as security multi-party computation, differential privacy and homomorphic encryption.

According to the distribution characteristics of data's, federated learning can be divided into three categories: (1) horizontal federated learning, which is suitable for scenarios that user data have same features space, but the sample space is very different, such as two shops which are situated in different countries but sell same goods. (2) vertical federated learning, which is suitable for the completely opposite scenarios to horizontal federated learning: data that share same sample space but different feature space such as a restaurant and a barber which are both on a same avenue so that their customers are same as well. (3) federated transfer learning, in conditions that neither the feature space nor the sample space of the data is same enough, think

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

about a bank in China and a chemical plant in Japan, they have little in common due to geographical location and user groups.

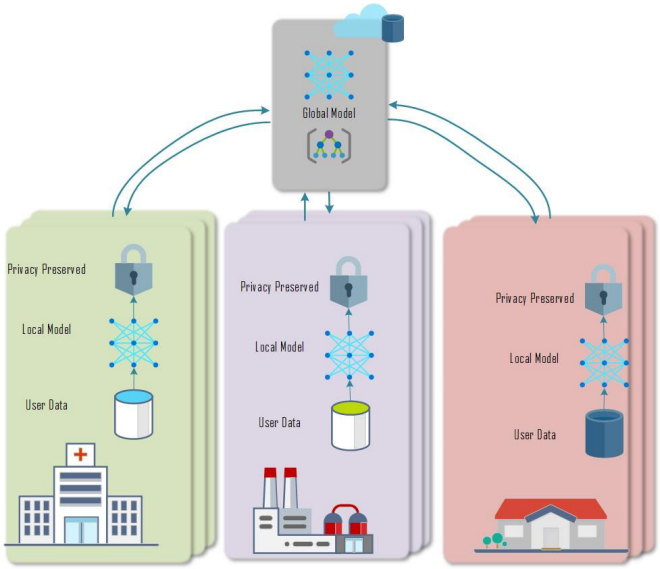

Figure 2. The framework of Federated learning

3. PROBLEM FORMULATION

3.1. System Model

The system model of this paper is shown in the Figure 3. In this model, it contains five participant entities, trusted authority, cloud/service provider, fog node, user and terminal equipment. These entities are described in detail below. Suppose there are n user entities that want to share the data model with other users instead of the data itself.

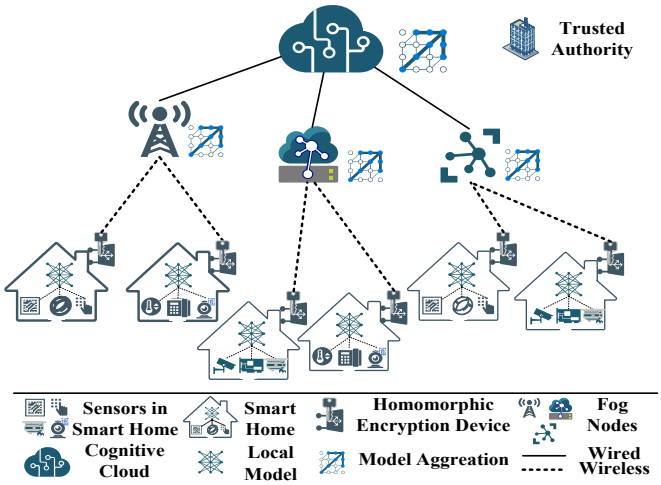

Figure 3. System model

1) Trusted authority (TA): TA is responsible for guiding the whole system, generating system security parameters and public and private keys. The TA is considered to be completely trusted.

2) Cloud/Service provider: The cloud can be seen as a service provider. The cloud is responsible for integrating the federated model and verifying the user's identity. But it can't get the original data and model parameters of users, so it is considered as semi-trusted.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

3) Fog node: The fog node is responsible for verifying user identity and aggregating user model parameters. The fog node is semi trusted and tries to obtain the user's model parameters.

4) User: Users are responsible for collecting terminal data, training local model and sending model parameters to fog nodes. Users are considered semi trusted and try to get model parameters of other users.

5) Terminal equipment: Terminal equipment is an edge terminal deployed on the user side, which can generate data through sensors. Terminal devices are completely trusted.

3.2. Security Requirements

The scheme proposed in this paper should meet the following security requirements. Trusted authority and terminal equipment are considered to be completely trusted. The cloud, fog node and user are considered semi trusted and try to get model parameters of the users. In addition, the model should be able to resist common malicious attacks, including but not limited to replay attacks, forgery attacks, malicious data injection attacks, denial of service attacks.

3.3. Design Goals

This paper proposes a new data sharing scheme, which needs to meet the following design objectives.

1) Confidentiality. The proposed scheme should satisfy confidentiality. User's local model parameters belong to user privacy data, so privacy disclosure should be prevented.

2) Authentication. The proposed scheme should meet the requirements of authentication. To prevent malicious data injection attacks or illegal access, the design should be able to identify illegal users and filter out malicious data.

3) Efficiency. The proposed scheme should meet the requirements of efficiency. Not only the model training should converge quickly, but also the encryption and signature protocols should have lower computational complexity and communication cost.

4) High accuracy and low loss. The training model of Federated learning designed in this paper should have higher accuracy and lower loss.

4. PROPOSED SCHEME

In this section, we will introduce the proposed scheme in detail, including system initialization, local training, report generation, model aggregation, report reading, global model and model update. The main symbols and descriptions of the proposed scheme are shown in Table 1.

Table 1. The Main Symbols and Descriptions of The Proposed Scheme

Symbol	Description
s	System master secret key
ι	System security parameters
$\mathbb{G}, \mathbb{G}_T$	Additive groups
p,q	Large prime number
g_1, g_2, g_3	Generators of $\mathbb{G}$
$Ppub$	System master public key
H_1, H_2	Hash functions
ID	Identification
PK_{ID}	The public key of ID
SK_{ID}	The private key of ID
PID_{ID}	The pseudonym of ID
$\oplus$	Bitwise XOR operation

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

4.1. System Initialization

$Setup$: Given a system security parameter ι as input, TA selects an elliptic curve $E(F_p)$ with order q, two cyclic groups $\mathbb{G}$ and $\mathbb{G}_T$, where p, q are two large prime numbers, a bilinear mapping $e: \mathbb{G} \to \mathbb{G}_T$ and generators g_1, g_2, g_3. Two secure hash functions: H_1: $\{0,1\}^* \to \mathbb{Z}_q^*$, and H_2: $\{0,1\}^* \to E(F_p)$. TA randomly selects system master key $MSK = s \in \mathbb{Z}_q^*$. TA public system parameters $params = \{E(F_p), \mathbb{G}, \mathbb{G}_T, \iota, e, p, q, g_1, g_2, g_3, H_1, H_2\}$.

$Keygen$: The encryption scheme proposed in this paper is based on the Paillier homomorphic encryption system [27]. The details of key generation are as follows: Given two large prime numbers p, q, calculate RSA module $n = pq$ and Carmichael function $\lambda = (p-1)(q-1)$. Suppose g_1 is an ordered n generator on $\mathbb{Z}_{n^2}^*$, satisfying ${g_1}^n \bmod n^2 = 1$. Define the function $L(u) = (u-1)/n$ and $\mu = 1/[L({g_1}^\lambda \bmod n^2)]$. Then the corresponding public and private key pairs are: $pk = (n, g_1)$ and $sk = (\lambda, u)$. TA transmits sk to cloud through secret channel.

TA computes the partial public key $PK_{i,1} = H_2(ID_i)$ for each system user (fog user or edge user) with ID_i, calculates the private key $SK_i = s \cdot PK_{i,1}$, the other partial public key $PK_{i,2} = SK_i \cdot g_1$ and transmits SK_i to them through secret channel. In addition, TA generates a pseudonym PID_i for each edge user U_i to upload and share data and transmits to U_i through secret channel. This pseudonym can be updated dynamically from time to time. The update method is as follows. In time t_0, U_i's pseudonym is PID_i. In time t_1, U_i gets its new pseudonym $PID_i' = PID_i \oplus SK_i \oplus t_1$.

4.2. Local Training

Suppose that users U_i want to share data with other users, U_i first initiate a sharing request to Cloud. Cloud sends sharing instructions to all users, where instructions consists of the initial model, optimizer, and timestamp. Other users choose to share or not to share according to their own wishes. If they choose not to share, the instruction will be discarded. Otherwise, local training and report generation are performed. Suppose n users agree to share data at time t.

Each user performs local training to get model LM_i^k (k is the current iteration index) and parameters $MP_i = \{w^i\} \in \mathbb{Z}_n$, where w^i is the weight. In consideration of local user's calculating ability and the need for time-effectiveness, the selection of local model is mainly shallow convolutional neural networks with lone short-term memory networks or gated recurrent neural networks, which is able to capture the temporal characteristics of network traffic data. Each user optimizes the local model LM_i^k at themselves data. Their goal is to minimize the local loss function $L(LM_i^k)$,

$$LM_i^k = L(w^i) \tag{1}$$

4.3. Report Generation

At time t, suppose the plaintext of U_i is $MP_i \in \mathbb{Z}_n$. The number r_i is randomly selected by U_i to satisfy the greatest common divisor $gcd(r_i, n) = 1$. Then the calculation method of ciphertext c_i is: $c_i = g_1^{MP_i} \cdot r_i^n \bmod n^2$. U_i calculates $h_i = H_1(c_i)$, generates the signature $\sigma_i = SK_i \cdot g_2 + SK_i \cdot h_i \cdot g_3$, and computes $v_i = H_1(PID_i \parallel c_i \parallel \sigma_i \parallel t)$. U_i generates the report $M_i = \{PID_i, c_i, \sigma_i, v_i, t\}$ and forward to fog f_j.

4.4. Model Aggregation

Inspired by FedAvg algorithm [23], we design the aggregation algorithm of the model, as shown in Fig. 3. After the fog f_j receives the message M_i from U_i, he gets $PID_i, c_i, \sigma_i, v_i, t$,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

checks timestamp t and data integrity. The method is to check whether $v_i' = H_1(PID_i \| c_i \| \sigma_i \| t) = v_i$ is true. If all user's data integrity is accepted, f_j verify the signatures of these users in bulk. Otherwise, $f_j$discard the invalid message. f_j calculates $h_i' = H_1(c_i)$, check whether $e(\sum_{i=1}^{n} \sigma_i, g_1) = e\left(g_2, \sum_{i=1}^{n} PK_{i,2}\right) \cdot e\left(g_3, \sum_{i=1}^{n} (PK_{i,2} \cdot h_i')\right)$ is true. If true, f_j perform the following model aggregation operation, otherwise, report errors and discard data.

f_j calculates the model aggregation $c_{agg} = \prod_{i=1}^{n} c_i$, $h_j = H_1(c_{agg})$, generates the signature $\sigma_j = SK_j \cdot g_2 + SK_j \cdot h_j \cdot g_3$, and computes $v_j = H_1(ID_j \| c_{agg} \| \sigma_j \| t)$. f_j generates the aggregation report $M_{agg} = \{ID_j, c_{agg}, \sigma_j, v_j, t\}$ and forward to cloud.

4.5. Report Reading

After the cloud receives the message M_{agg} from f_j, he checks timestamp t and data integrity by check whether $v_j' = H_1(ID_j \| c_{agg} \| \sigma_j \| t) = v_j$ is true. Once accepted, the cloud checks the f_j's signature σ_j whether is valid. The cloud calculates $h_j' = H_1(c_{agg})$, checks whether $e(\sigma_j, g_1) = e(g_2, PK_{j,2}) \cdot e(g_3, PK_{j,2} \cdot h_j')$ is true. If so, cloud recognizes the message and performs the decryption operation. Otherwise, discard the invalid message.

Suppose the ciphertext is $c_{agg} \in \mathbb{Z}_{n^2}^*$, then the calculation method of plaintext M_{agg} is: $M_{agg} = (L(c_{agg}^{\lambda} \bmod n^2)/L(g_1^{\lambda} \bmod n^2) \bmod n)$.

4.6. Global Model

After the cloud gets the message $M_{agg} = \sum_{i=1}^{n} MP_i = \sum_{i=1}^{n} weight_i$ by decryption operation, he sets $weight = \sum_{i=1}^{n} weight_i$. Cloud updates the model parameters and gets the global model GM_G^k with $weight$. Cloud broadcast the global model GM_G^k to all edge users.

$$GM_G^k = \frac{1}{n} \sum_{i=1}^{n} LM_i^k \tag{2}$$

4.7. Personality Choice

After U_i receives the global model GM, he can choose whether to update according to his own will, based on the accuracy and loss of model training LM_i and GM.

4.8. Model Update

If U_i chooses to update his model using the global model GM, he can replace the original model with the global model GM as the initial model of the next round of training.

5. SECURITY ANALYSIS

5.1. Planning

This section considers the security analysis, as following.

1) Confidentiality. Paillier homomorphic encryption system is based on the deterministic composite residual problem. It has been proved that it is semantically secure and can effectively resist chosen plaintext attack [27].

2) Authentication. The authentication of this scheme is guaranteed by the signature protocol. The security of the signature protocol in this paper has been proved [28], which ensures the inaccessibility of illegal users and the filtering of malicious data.

3) Replay attacks. Replay attack uses the time difference of data transmission to cover up the attack. In this paper, U_i computes $v_i = H_1(PID_i \| c_i \| \sigma_i \| t)$ and forward to fog f_j, where t is the current time. f_j checks whether $v_i' = H_1(PID_i \| c_i \| \sigma_i \| t) = v_i$ is true or not. If a malicious user performs a replay attack, $v_i'' \neq v_i$ and the attack is invalid.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

4) Forgery attacks. Forgery attack is that malicious or illegal users engage in illegal activities by forging the identity of legal users. According to the above authentication, the signature protocol is secure, so the forgery attack will be recognized.

5) Malicious data injection attacks. Malicious data injection attacks disrupt the normal operation of the system by injecting a large amount of malicious data. However, the design principle of this scheme is to verify the user's identity first. If there is no legal identity, the data will be recognized and discarded. Malicious data injection attack is invalid.

6) Denial of service attacks. Denial of service (DOS) attack is a malicious attack that causes the system to fail to run normally by launching a large number of service requests. In this scheme, the way to resist denial of service attack is authentication. Therefore, this attack is invalid in this article.

6. EXPERIMENT AND PERFORMANCE ANALYSIS

The scheme proposed in this paper can achieve efficiency, high accuracy and low loss. The digital simulation experiments are as follows.

6.1. Data Set

The UNSW-NB15 [29] dataset is used to verify our scheme's performance. In consideration of the difference of each local terminal, the data that is used to train local model is a set of subgroups of the dataset which have some intersections. And each local terminal is allocated with a parameter to decide the proportion of its model during the global model aggregation, which depends on the size of local data subset.

6.2. Federal Learning and Training

We implement CNN with LSTM on UNSW-NB15 datasets and set model dropout is 0.1. The detailed training results are shown in the Figure. 4. In the Figure. 4 (left), taking two users as an example, the result represents the result of local training for a single user. In the Fig. 4(right), it represents the training result of another user based on the model after the model parameters are shared. From the results, model sharing helps to improve the accuracy of model training, accompanied by the decline of loss.

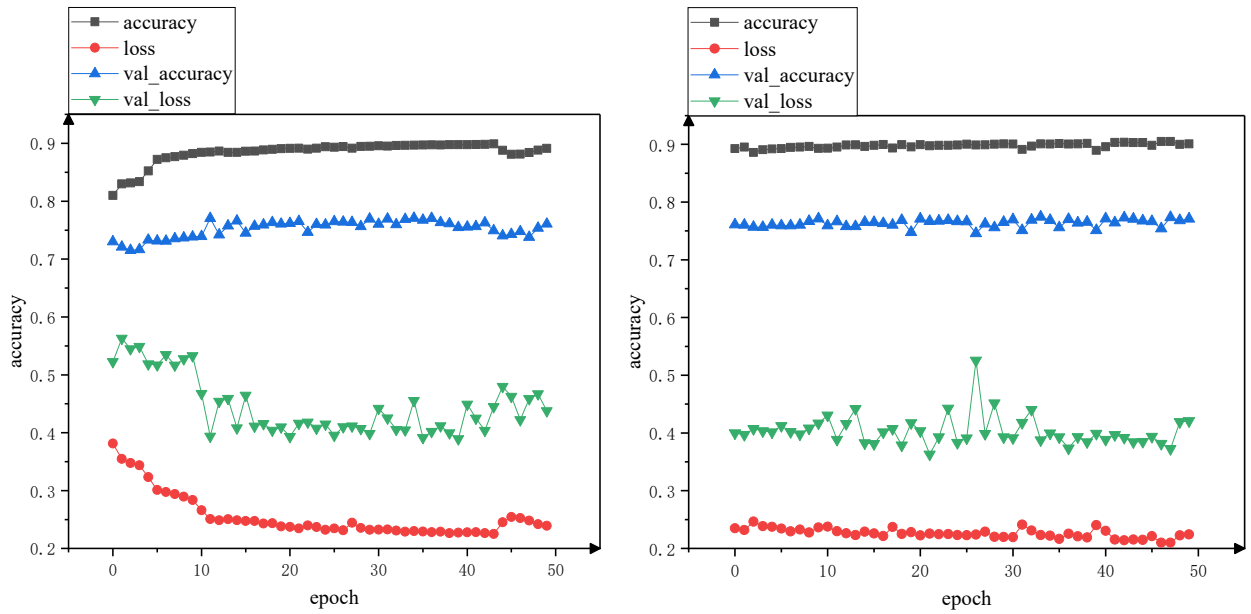

Figure 4. The detailed training results of FL.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

6.3. Digital Simulation

Digital simulation experiments are carried out to verify the performance of the proposed scheme, including computational complexity and communication cost. The simulation environment is Intel(R) Core (TM) i5-10210U CPU @1.6GHz 2.11GHz, RAM 8G and JPBC 2.0 library. In this paper, T_p, T_h, T_e and T_m mean the time of a paring operation, a hash operation, an exponent operation and a scalar multiplication operation, respectively. In order to reduce the random error, each experimental result is averaged 1000 times. $|\mathbb{G}|$ and $|m|$ mean the length of values of $\mathbb{G}$ and the message that has sent, respectively. In here, we consider three phases, including report generation, model aggregation and report reading.

In the report generation phase, U_i computes $c_i = g_1^{MP_i} \cdot r_i^n \bmod n^2$, calculates $h_i = H_1(c_i)$, generates the signature $\sigma_i = SK_i \cdot g_2 + SK_i \cdot h_i \cdot g_3$, and computes $v_i = H_1(PID_i \parallel c_i \parallel \sigma_i \parallel t)$. The user needs to perform two hash operation, four scalar multiplication operation and two exponent operation. U_i forwards $M_i = \{PID_i, c_i, \sigma_i, v_i, t\}$ to fog f_j. In the same way, there need $(2n+2)$ hash operation, $(2n+3)$ scalar multiplication operation and three paring operation in the model aggregation phase. In the report reading phase, there need two hash operation, three paring operation and two exponent operation as shown in Table 2. Performance analysis of the proposed scheme is shown in Figure.5 In the figure, the four lines from top to bottom represent the computational complexity of the whole scheme, the computational complexity of the Report generation phase, the computational complexity of the Model aggregation phase and the computational complexity of the Report reading phase, respectively.

Table 2. Performance Analysis of The Proposed Scheme

Phase	Computational complexity	Communication cost
Report generation	$\boldsymbol{n(2T_h + 4T_m + 2T_e)}$	$\boldsymbol{n(\|\mathbb{G}\| + 4\|m\|)}$
Model aggregation	$\boldsymbol{(2n+2)T_h + 3T_p + (2n+3)T_m}$	$\boldsymbol{\|\mathbb{G}\| + 4\|m\|}$
Report reading	$\boldsymbol{2T_h + 3T_p + 2T_e}$	$\boldsymbol{\|m\|}$

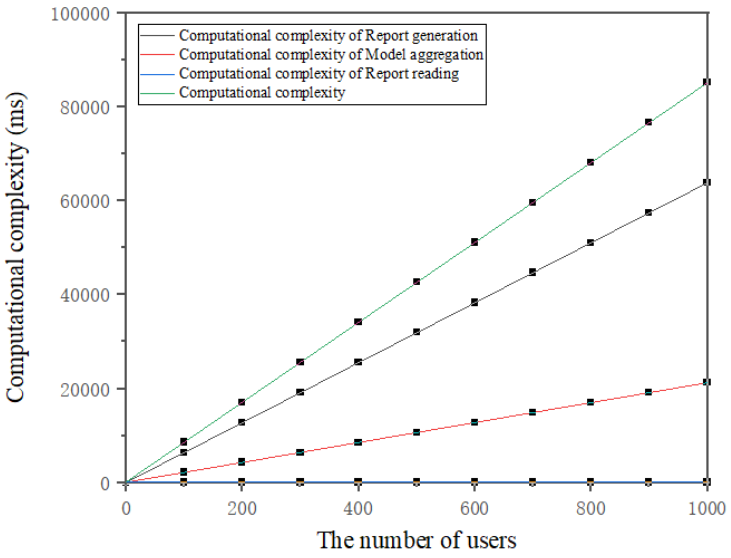

Figure 5. Performance analysis of the proposed scheme.

7. RELATED WORKS

Cognitive Internet of Things was proposed by Wu et al. in 2014 [14]. Since it was put forward, this concept has drawn wide attention. In 2017, to solve the problem of achieving the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

appropriate strategy to transmit packets from different buffers through multiple channels to maximize the system throughput, a new Q-learning-based (deep learning) transmission scheduling model for the CIoT is proposed [15]. In 2018, Ding et al. surveyed to show the state-of-the-art studies on amateur drone surveillance [16]. They proposed a cognitive Internet of Things framework for amateur drone surveillance. In 2019, Li et al. investigated the structural frameworks and potential applications of cognitive IoT [17]. In 2019, Liu et al. proposed a new metric, called the Quality of Information Coverage (QIC), which characterizes information coverage quality and rewards for data sensing to maximize the QIC [18].

Yang et al. in 2017 outlined an architecture for a people-centric and cognitive Internet of Things (PIoT) environmental sensing platform, which involves closed loops of interactions among people nodes and physical devices as well as servers and recommendations on device connections by cognitive computing [19]. In 2020, Liu et al. proposed a non-orthogonal multiple access (NOMA)-based hybrid spectrum access scheme for 6G-enabled cognitive IoT (CIoT), where the CIoT may access both the idle and busy spectrum via NOMA regardless of the PU's state [20].

Federated learning was first proposed by Konečný et al. in 2016 to solve the problem of mass terminal mobile phone system update [21]. Since federal learning was proposed, it has attracted a large number of researchers' attention because of its subversive data sharing mode [22]. In 2020, Chen et al. studied the problem of training federated learning (FL) algorithms over a realistic wireless network [23]. In the considered model, wireless users execute an FL algorithm while training their local FL models using their own data and transmitting the trained local FL models to a base station (BS) that generates a global FL model and sends the model back to the users.

In 2020, Niknam et al. provided an accessible introduction to the general idea of federated learning, discuss several possible applications in 5G networks, and describe key technical challenges and open problems for future research on federated learning in the context of wireless communications [24]. Karimireddy et al. in 2020 proposed a new algorithm (SCAFFOLD) which uses control variates (variance reduction) to correct for the 'client drift' [25]. In 2020, Khan et al. presented the primary design aspects for enabling federated learning at the network edge [26]. In 2021, Li et al. to protect user privacy in smart grid from the perspective of blockchain [30]. The comparison between different schemes is shown in Table 3.

Table 3. The Comparison of Different Schemes

Schemes	Anomaly Detection	FL Model aggregation	Confidentiality	Authentication	Lightweight
[9]	√	×	×	×	×
[10]	√	×	×	×	×
[11]	√	×	×	×	√
[12]	√	×	×	×	√
[13]	√	×	×	×	×
[23]	√	√	×	×	√
ours	√	√	√	√	√

8. CONCLUSION

In this paper, we propose an anomaly detection method based on federated learning and homomorphic encryption on cognitive Internet of Things for fog-based smart home. Firstly, a model based on federated learning on cognitive Internet of Things for fog-based smart home is proposed to solve the problem of data island through multi model cooperation of different

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

terminals. Secondly, to protect the model parameters and other user privacy data, we use homomorphic encryption to achieve encrypted transmission. By using homomorphic encryption, data aggregation of model parameters is realized. After analysis, the model can be extended to multi-dimensional or multi-level. Finally, security analysis, model training performance, computational complexity and communication cost are investigated. The simulation results show that the proposed scheme has good performance and low cost. However, there is still room for improvement in the federal learning algorithm, which is also the next work plan.

ACKNOWLEDGMENT

Thanks to editors and anonymous peer reviewers.

REFERENCES

[1] Bonomi, F., Milito, R., Zhu, J., & Addepalli, S, "Fog computing and its role in the internet of things," In Proceedings of the first edition of the MCC workshop on Mobile cloud computing, pp. 13-16, August 2012.

[2] Yi, S., Li, C., & Li, Q, "A survey of fog computing: concepts, applications and issues," In Proceedings of the 2015 workshop on mobile big data, pp. 37-42, June 2015.

[3] Yi, S., Hao, Z., Qin, Z., & Li, Q, "Fog computing: Platform and applications," In 2015 Third IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb), pp. 73-78, IEEE, November 2015.

[4] Vaquero, L. M., & Rodero-Merino, L, "Finding your way in the fog: Towards a comprehensive definition of fog computing," ACM SIGCOMM Computer Communication Review, vol. 44, no. 5, pp. 27-32, 2014.

[5] Bansal, M., Chana, I., & Clarke, S, "Enablement of IoT based context-aware smart home with fog computing," Journal of Cases on Information Technology (JCIT), vol. 19, no. 4, pp. 1-12, 2017.

[6] Verma, P., & Sood, S. K, "Fog assisted-IoT enabled patient health monitoring in smart homes," IEEE Internet of Things Journal, vol. 5, no. 3, pp. 1789-1796, 2018.

[7] Rahimi, M., Songhorabadi, M., & Kashani, M. H, "Fog-based smart homes: A systematic review," Journal of Network and Computer Applications, pp. 153-153, 2020.

[8] Bhatia, M, "Fog Computing-inspired Smart Home Framework for Predictive Veterinary Healthcare," Microprocessors and Microsystems, pp. 78-78, 2020.

[9] Chandola, V., Banerjee, A., & Kumar, V, "Anomaly detection: A survey," ACM computing surveys (CSUR), vol. 41, no. 3, pp. 1-58, 2009.

[10] Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K, "Network anomaly detection: methods, systems and tools," IEEE communications surveys & tutorials, vol. 16, no. 1, pp. 303-336, 2013.

[11] Nachman, B., & Shih, D, "Anomaly detection with density estimation," Physical Review D, vol. 101, no. 7, pp. 075042, 2020.

[12] Fei, Y., Huang, C., Jinkun, C., Li, M., Zhang, Y., & Lu, C, "Attribute restoration framework for anomaly detection," IEEE Transactions on Multimedia, 2020.

[13] Garg, S., Kaur, K., Batra, S., Kaddoum, G., Kumar, N., & Boukerche, A, "A multi-stage anomaly detection scheme for augmenting the security in IoT-enabled applications," Future Generation Computer Systems, vol. 104, pp. 105-118, 2020.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202211_8(11).0026

[14] Wu, Q., Ding, G., Xu, Y., Feng, S., Du, Z., Wang, J., & Long, K, "Cognitive internet of things: a new paradigm beyond connection," IEEE Internet of Things Journal, vol. 1, no. 2, pp. 129-143, 2014.

[15] Zhu, J., Song, Y., Jiang, D., & Song, H, "A new deep-Q-learning-based transmission scheduling mechanism for the cognitive Internet of Things," IEEE Internet of Things Journal, vol. 5, no. 4, pp. 2375-2385, 2017.

[16] Ding, G., Wu, Q., Zhang, L., Lin, Y., Tsiftsis, T. A., & Yao, Y. D, "An amateur drone surveillance system based on the cognitive Internet of Things," IEEE Communications Magazine, vol. 56, no. 1, pp. 29-35, 2018.

[17] Li, F., Lam, K. Y., Li, X., Sheng, Z., Hua, J., & Wang, L, "Advances and emerging challenges in cognitive internet-of-things," IEEE Transactions on Industrial Informatics, vol. 16, no. 8, pp. 5489-5496, 2019.

[18] Liu, Y., Liu, A., Wang, T., Liu, X., & Xiong, N. N, "An intelligent incentive mechanism for coverage of data collection in cognitive Internet of Things," Future Generation Computer Systems, vol. 100, pp. 701-714, 2019.

[19] Yang, L., Li, W., Ghandehari, M., & Fortino, G, "People-centric cognitive internet of things for the quantitative analysis of environmental exposure," IEEE Internet of Things Journal, vol. 5, no. 4, pp. 2353-2366, 2017.

[20] Liu, X., Ding, H., & Hu, S, "Uplink Resource Allocation for NOMA-based Hybrid Spectrum Access in 6G-enabled Cognitive Internet of Things," IEEE Internet of Things Journal, 2020.

[21] Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D, "Federated learning: Strategies for improving communication efficiency," arXiv preprint arXiv:1610.05492, 2016.

[22] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V, "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50-60, 2020.

[23] Chen, M., Yang, Z., Saad, W., Yin, C., Poor, H. V., & Cui, S, "A joint learning and communications framework for federated learning over wireless networks," IEEE Transactions on Wireless Communications, 2020.

[24] Niknam, S., Dhillon, H. S., & Reed, J. H, "Federated learning for wireless communications: Motivation, opportunities, and challenges," IEEE Communications Magazine, vol. 58, no. 6, pp. 46-51, 2020.

[25] Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T, "Scaffold: Stochastic controlled averaging for federated learning," In International Conference on Machine Learning, pp. 5132-5143, PMLR, November 2020.

[26] Khan, L. U., Pandey, S. R., Tran, N. H., Saad, W., Han, Z., Nguyen, M. N., & Hong, C. S, "Federated learning for edge networks: Resource optimization and incentive mechanism," IEEE Communications Magazine, vol. 58, no. 10, pp. 88-93, 2020.

[27] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in Proc of EUROCRYPT'99, Czech Republic, May, pp.223–238, 1999.

[28] Liu, Y., Guo, W., Fan, C. I., Chang, L., & Cheng, C, "A practical privacy-preserving data aggregation (3PDA) scheme for smart grid," IEEE Transactions on Industrial Informatics, vol. 15, no. 3, pp. 1767-1774, 2018.

[29] Moustafa, Nour, and Jill Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," 2015 military communications and information systems conference (MilCIS), IEEE, pp. 1-6, 2015.

[30] Li, K., Yang, Y., Wang, S., Shi, R., & Li, J., "A lightweight privacy-preserving and sharing scheme with dual-blockchain for intelligent pricing system of smart grid," Computers & Security, 103, 102189, 2021.