Egyptian Informatics Journal 24 (2023) 217–227

Contents lists available at ScienceDirect

Egyptian Informatics Journal

journal homepage: www.sciencedirect.com

Research on smart city data encryption and communication efficiency improvement under federated learning framework

Zhen Kuang [a,b], Chaoyang Chen [b,*]

[a] Faculty of Data Science, City University of Macau, Macau, Macau, China
[b] School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, Hunan, China

ARTICLE INFO

Article history:
Received 19 January 2023
Revised 23 February 2023
Accepted 27 February 2023
Available online 23 March 2023

Keywords:
Federal learning
Smart city
Privacy security
Formula encryption
Edge computing
Asynchronous communication
Differential privacy

ABSTRACT

To improve the data communication processing capability of smart city informatization construction and defend against malicious joint attacks of internal communication participants, we studied the use of function encryption, blockchain, differential privacy and other technologies to defend against weight disclosure, participation in collusion attacks, single point failures and other issues in the federal learning process, and introduced edge computing and asynchronous communication in the classic federal learning framework, Improve the communication efficiency of the model while ensuring the accuracy. The research results show that the accuracy of FE-BDP algorithm can maintain above 95% for a long time, and the change is not significant with the increase of the number of users, indicating that the algorithm has strong stability. The loss value of FE-LDP model is significantly smaller than that of other models and can be stabilized below 0.05. The data aggregation time of blockchain encryption technology is less than 1.0s. The edge-asynchronous communication framework can be applied to multiple urban scenarios and achieve effective data communication, with the highest accuracy rate of 93.87%and the lowest communication cost of 1315.29s. The research results show that the security encryption fusion technology can effectively protect user data privacy, and the edge-asynchronous communication framework has obvious effect on improving communication efficiency, which has important application value for promoting the construction of smart city informatization.

1. Introduction

With the growth of China's population, the urban population also displays a trend of ceaseless growth, which leads to the increasingly difficult urbanization governance. In this context, the proposal of smart cities tenders ideas for urban governance. The fundamental idea of smart city is to use Internet technology to realize the digitalization of urban traffic, water, electricity and other data, so as to facilitate the daily life of urban residents and the unified management of the city [1,2]. A large number of personal and enterprise data have been stored in Internet big data, and their data security has begun to attract the attention of the public. However, the traditional data encryption and communication techniques are laborious to achieve the security management of large-dimension urban data, and the probability of data leak will still occur [3]. For this reason, research has advanced a Federated Learning (FL) framework, which solves the privacy preservation trouble by building a FL pattern, and can avoid customer information disclosure while making full use of multi-party data [4,5]. However, with the ceaseless expansion of the dimension of urban data, the encryption difficulty of urban data is also increasing, and the traditional FL has been laborious to achieve effective preservation of customer data [6]. For this reason, the research has introduced a differential privacy preservation technology under the FL to further realize the encryption of smart city data. It has introduced the Block chain (BC) to address the trouble of urban privacy data leak caused by failures. Then, it uses the edge asynchronous framework to achieve effective data transmittal and ameliorate the productivity of data communication, so as to tender a theoretical cornerstone for the development of smart cities.

* Corresponding author.
E-mail address: cychen@hnust.edu.cn (C. Chen).

Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.eij.2023.02.005

2. Related works

The amount of data in modern information networks has shown an explosive growth trend, and the threat to subscribers' privacy security has become increasingly prominent. In the era of global information explosion, the amount of information network data has also displayed an explosive growth trend, and the threat to subscribers' privacy security has become increasingly prominent. To ameliorate the security of data interaction and sharing, many researchers have conducted research on privacy preservation and security encryption. Xi P and other scholars use BC based FL to achieve the safe reposition and sharing of medical data. Taking advantage of BC's advantages in decentralization and attack prevention, they combine BC technology with the FL to establish a medical data sharing pattern, effectively improving the security and traceability in the process of medical data reposition and management [7]. To address the trouble of data islands and security threats in the social Internet of Things, Yin L advances a hybrid privacy preservation FL framework, which uses advanced formula encryption to securely encrypt the data and weights transmitted by customers, and introduces sparse differential gradient to ameliorate the transmittal productivity of the FL framework. The research outcomes display that the hybrid privacy preservation pattern has high transmittal exactitude and productivity [8]. Ou W and his team members advanced a vertical FL framework based on homomorphic encryption and Bayesian machine learning to solve the contradiction between data sharing and privacy preservation. In the vertical FL framework, the local data of the customer will be encrypted before data sharing, which can advance the sharing of subscriber information and ameliorate the security of subscriber privacy. The research outcomes display that the exactitude rate of the pattern under the vertical FL is 90%, which can be widely used in many fields such as medical education [9]. Arachchige P C M and other scholars introduced the FL in the industrial IoT, combined the FL with machine learning, differential privacy and BC technology, and advanced a framework called PriModChain to ameliorate the privacy, security and reliability of industrial IoT data [10]. Directing the troubles of data sharing and information security in the industrial Internet of Things, the Jia B team advanced a FL based on BC, which combines BC technology with the FL to achieve data security aggregation. In addition, Jia B team also combined differential privacy with homomorphic encryption and random forest algorithm to further defend the privacy data in the data sharing pattern. A large number of experimental outcomes have proved that the aggregation scheme has good performance in the industrial Internet [11]. Yu R introduced mobile edge computing into the FL to optimize the resources of the FL and advance the efficient use of federal learning resources. Through the review of data compensation and hierarchical aggregation technology in FL, they advanced a neural perceptual resource management technology based on modular FL, which allocates global pattern subnets according to the resource status of mobile customers. The research outcomes display that the resource management pattern has high application flexibility and resource utilization [12]. Communication productivity is also a key trouble in the domain of data interaction and processing. Sun H advanced an optimization technique based on gradient sparsity for the communication productivity in the FL framework. Through gradient correction and local gradient and batch normalization, the gradient parameters are updated to reduce the impact of delay gradient on pattern calculation and reduce the communication overhead during pattern calculation, Advance better convergence of the pattern. The research carried out experimental verification on multiple datasets, indicating that the pattern has the highest exactitude at 99.9% sparse gradient [13]. Duan M advances a self balanced FL framework for unbalanced distributed data. It solves the trouble of data imbalance through data expansion and multi customer rescheduling, and uses adaptive data expansion and intermediaries to achieve global and local data equalization. The research outcomes display that the top-1 exactitude of the self balanced FL framework on the two datasets has been ameliorated by 4.39% and 6.51% respectively, and the communication productivity has been ameliorated by 75% [14]. For wireless network data communication, Zhong R advances a FL framework based on mobile reconfigurable intelligent surface, which combines the mobility advantage of reconfigurable intelligent surface with the FL framework to advance the gain of pattern data rate. The research outcomes display that the data rate gain of the FL framework reaches 42%, which has good optimization performance and convergence productivity [15]. To address the troubles of radio resource utilization productivity and energy productivity, Kaur A and other scholars advanced a multi-agent reinforcement learning scheme, which uses decentralized cooperation to randomly form a dynamic team of multi-agent to ameliorate the overall resource allocation productivity. The research outcomes display that the multi-agent reinforcement learning scheme has a fast convergence velocity, and has a significant effect in improving the network capacity, which can ameliorate the service quality and energy utilization productivity of subscribers [16]. To sum up, the FL is a frequently-used instruments of data communication and processing, but the previous federal learning security preservation technology still has privacy preservation loopholes, and data communication productivity and processing performance need to be ameliorated. Therefore, directing the security encryption and communication productivity under the FL, the research advances a security encryption fusion technology based on formula encryption, BC and other means, and designs an edge asynchronous communication framework, hoping to tender technical help for massive data processing and privacy security preservation in the informatization construction of smart cities.

3. Research on data aggregation encryption under FL

3.1. Data aggregation weight hidden protection

In the context of smart city informatization construction, in order to improve the security and communication efficiency of data communication under the federal learning framework, we studied the use of function encryption and Bayesian Differential Privacy (FE-BDP) model to protect the data aggregation weight, The decentralized key security management is realized by using function encryption and local differential privacy (FE-LDP) model. Aiming at the server attack and single point of failure under the federated learning framework, the research further introduces the function encryption and blockchain (FE-BC) model to achieve data security aggregation. In order to reduce the communication overhead of the federated learning framework and improve the efficiency of data communication, an edge asynchronous efficient communication framework combining edge computing and asynchronous communication is proposed. The data encryption and communication efficiency improvement method under the federated learning framework is shown in Fig. 1.

When subscribers share and communicate data in the FL (FL) framework, they are vulnerable to inversion attacks, reconstruction attacks and other privacy attacks, which seriously threaten subscribers' personal data security [17]. Homomorphic encryption, differential privacy and other technologies are commonly used to deal with the trouble of subscriber privacy leak. Communication data is encrypted by introducing noise into the communication signal [18,19]. However, common privacy preservation technologies

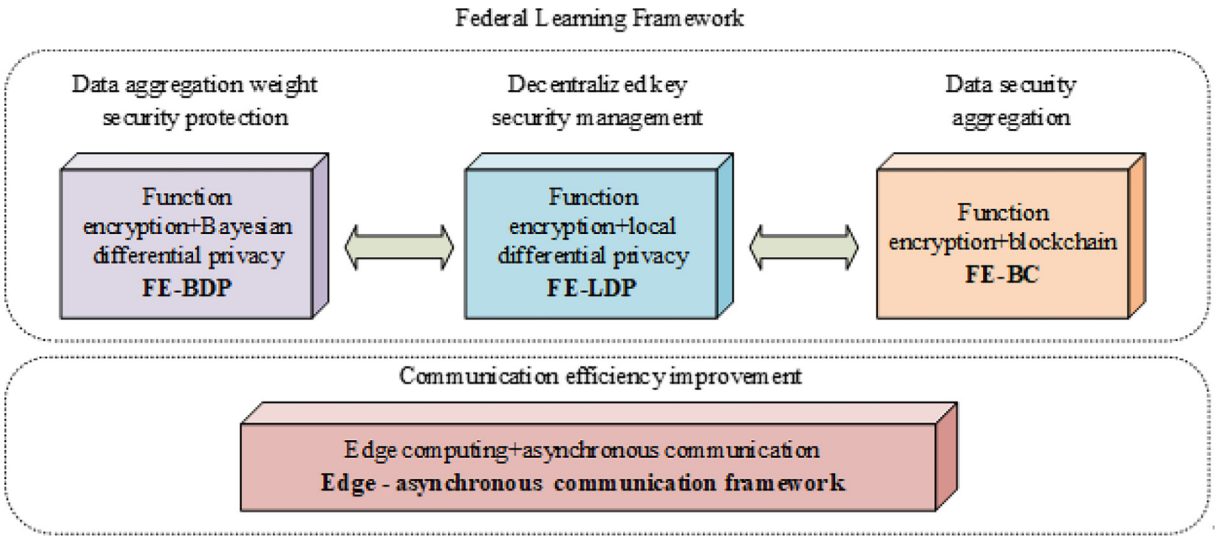

Fig. 1. Data encryption and communication efficiency improvement method under federated learning framework.

do not consider the security of subscribers' data aggregation weight, and malicious intruders may steal subscribers' native message by tampering with the weight. Therefore, to address the trouble of weight leak of data aggregation under the federated learning framework, we use Formula Encryption and Bayesian Differential Privacy (FE-BDP) technology to build a data aggregation weight security preservation pattern. The server cannot gain the weight parameters of other subscribers by using formulaal encryption, and introduces noise into the local pattern, and judges the privacy loss of the pattern by disturbing the distribution distance of the pattern, To ensure that the safety is ameliorated and the amount of noise introduced is reduced as much as possible, so as to avoid affecting the exactitude of the pattern due to excessive noise. The data aggregation weight security preservation pattern based on FE-BDP is displayed in Fig. 2.

Set the subscriber set as $\{C_1, C_2, \cdots, C_n\}$, there are n subscribers in total, and each subscriber's data set is $\{D_1, D_2, \cdots, D_n\}$. The key management party will generate and manage the key, and the subscriber's pattern parameters are polymerized in the server, and the pattern parameters are W_i. In the pattern training phase, formulaal encryption is used to encrypt the subscriber parameters. The subscriber sends his weighty$_i$ to the key manager, who combines the weights to gain vector and divides it into $y_1 \| \cdots \| y_n$, thus generating the formula decryption key$sk_{y_1 \| \cdots \| y_n}$. The key management party generates the major public-key mpk_i and the major secret-key msk_i , sends the encryption and decryption keys to the subscriber and the server, gains the subscriber's keys mpk_i and msk_i in combination with the subscriber tag$i$, synthesizes $pk_i := (mpk_i', msk_i)$ as the subscriber's public-key, uses pk_i to encrypt the pattern parameters, and sends the ciphertext c t_i^{out} to the server. The server decrypts the ciphertext to get $[a_i]_T$, and the weight aggregation pattern of all subscribers is the discrete logarithm of $[a_i]_T$. In the pattern prediction phase, Bayesian differential privacy is used to calculate the privacy loss in FL iterations. Suppose that the privacy loss in each iteration is $ct(\lambda)$, the data distribution is, and the expected value is $E\left[e^{\lambda D_{k+1}(p_i \| q_i)}\right]$. The privacy loss calculation formula is displayed in Eq. (1).

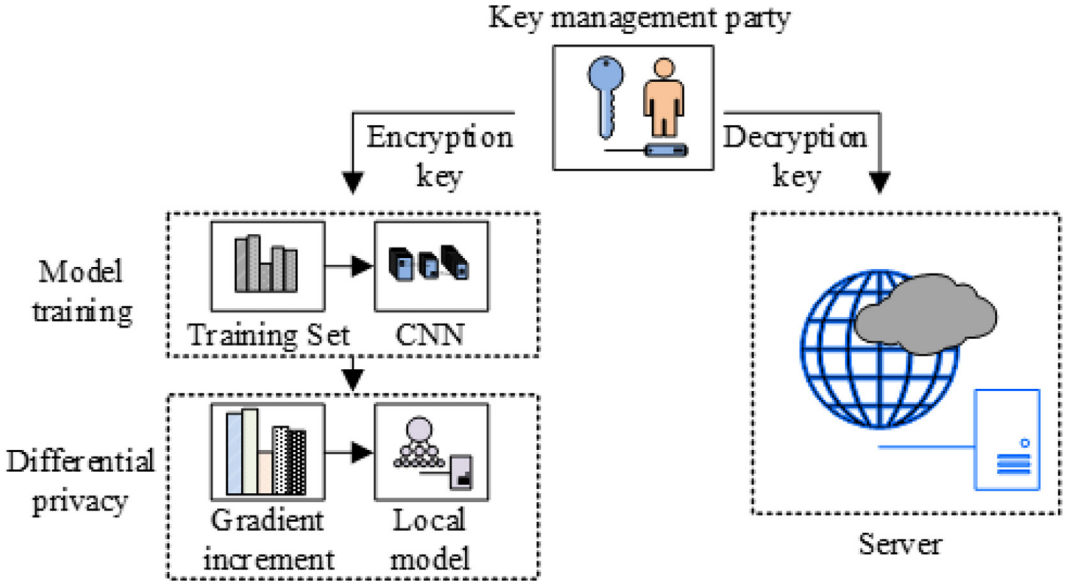

Fig. 2. Data aggregation weight security preservation pattern.

$$ct(\lambda) = \log E_x * \sum_{k=0}^{\lambda+1} \binom{\lambda+1}{k} q^k * (1-q)^{\lambda+1-k} e^{\frac{k^2-k}{2\sigma^2}|g_i - g_i|^2} \quad (1)$$

In Eq. (1), q is the probability of data sampling, g_t is the non private output of the t iteration with parameter x and without parameter x, σ is the noise parameter in the sub sampled Gaussian noise mechanism, λ is the quantity of experiments, and k is the quantity of experiments. Set privacy budget (PB) for privacy loss, and the budget conditions for privacy loss are displayed in Eq. (2).

$$\log \delta \leqslant \sum_{t=1}^{T} c_t(\lambda) - \lambda\varepsilon \quad (2)$$

FL requires multiple iterations of training, but the subscriber's parameters do not change much in adjacent iterations, so the subscriber will repeatedly send repeated parameters to the server in iterations, outcomeing in reduced training productivity of FL [20]. Therefore, the research adjusts the random gradient descent algorithm in traditional FL to a sparse differential gradient. After the first upload of gradient parameters, only the increment of parameters is transmitted to the server, and the threshold value of parameter increment is set. Parameters whose increment is lower than the threshold value will not be uploaded, so as to reduce the communication and encryption costs of parameter transmittal.

3.2. Decentralized key security management

To avoid the situation that the key manager and the server conspire to steal the subscriber's privacy information, the research advances an excitation mechanism based on local differential privacy (LDP) based on the concept of decentralization, and encrypts the subscriber's key formula, so that other agents cannot gain the subscriber's key. The decentralized key security management pattern based on formula encryption and local differential privacy (FE-LDP) is displayed in Fig. 3. The subscriber generates personal encryption-key ek_i by combining the major public-key and native message, encrypts the local pattern to gain encryption pattern $[c_i]_1$, and interacts with each other to gain parameters T_i and $\sum_i^n T_i = 0$. The subscriber generates personal secret-key sk_i by combining the encryption-key and T_i. The subscriber uploads a partial decryption key and an encrypted local pattern to the server. The partial decryption key $dk_{\bar{y},i}$ is generated by combining the secret-key sk_i and the weight $\bar{y}$ in the interior product formula. The server combines the partial decryption keys of all subscribers to gain the formulaal decryption key $dk_{\bar{y}}$, uses $dk_{\bar{y}}$ to achieve information security aggregation without decrypting the subscriber pattern, and does not rely on the key management party to generate and manage the key, so as to avoid the trouble of conspiring to steal subscriber information. The LDP based on Gaussian noise is used to train the subscriber's local pattern, so that subscribers can determine the PB and noise parameters in the light of their privacy needs. The calculation formula of privacy loss $c(o; M, aux, d, d')$ is displayed in Eq. (3).

$$c(o; M, aux, d, d') \triangleq \log \frac{pr[M(aux, d) = 0]}{M(aux, d') = 0} \quad (3)$$

In Eq. (3), d, d' is adjacent data sets, M is the random algorithm in LDP, o is the output of the random algorithm, and aux is the auxiliary input. All steps of privacy loss are the logarithm $a_M(\lambda; aux, d, d')$of the moment generating formula, and the definition is displayed in Eq. (4).

$$a_M(\lambda) \triangleq \max_{aux,d,d} a_M(\lambda; aux, d, d') E_{o\sim M(axx,d)} * \left[\exp\left(\lambda c(o; M, aux, d, d')\right)\right] \quad (4)$$

The maximum value of $a_M(\lambda; aux, d, d')$ is $\max_{aux,d,d} a_M(\lambda; aux, d, d')$. For any λ_ε, $\varepsilon > 0$, the threshold value of privacy loss is displayed in Eq. (5).

$$\delta = \min_{\lambda} \exp\left(a_M(\lambda) - \lambda\varepsilon\right) \quad (5)$$

When the loss of privacy exceeds the threshold, it means that the risk of privacy leak is high and the training needs to be stopped immediately. LDP allows subscribers to decide PB independently, but subscribers usually choose a high PB, which reduces the exactitude of the pattern [21]. Therefore, the study introduces an excitation mechanism. If the lower limit of subscriber privacy preservation acceptance is $\underline{pri}$, then the subscriber privacy preservation meets conditionpri(eps, sigma) $\geqslant \underline{pri}$, where eps is the PB and $sigma$ is the noise parameter. The server sends the public test data set and initial pattern to the subscriber in the iteration process. The subscriber upload the local pattern rate as a competitive tender to the server. The server will incorporate the subscribers with high exactitude into the federal learning and reward them. Under the excitation mechanism, subscribers and servers will try their best to pursue the balance between exactitude and privacy preservation, so as to find appropriate preservation strategies.

3.3. Data aggregation based on BC

In the FL framework, the server is an important node for subscriber data aggregation, but the server is vulnerable to attacks and single point of failure [22,23]. Therefore, the research introduces BC technology into classical FL, and uses the internal

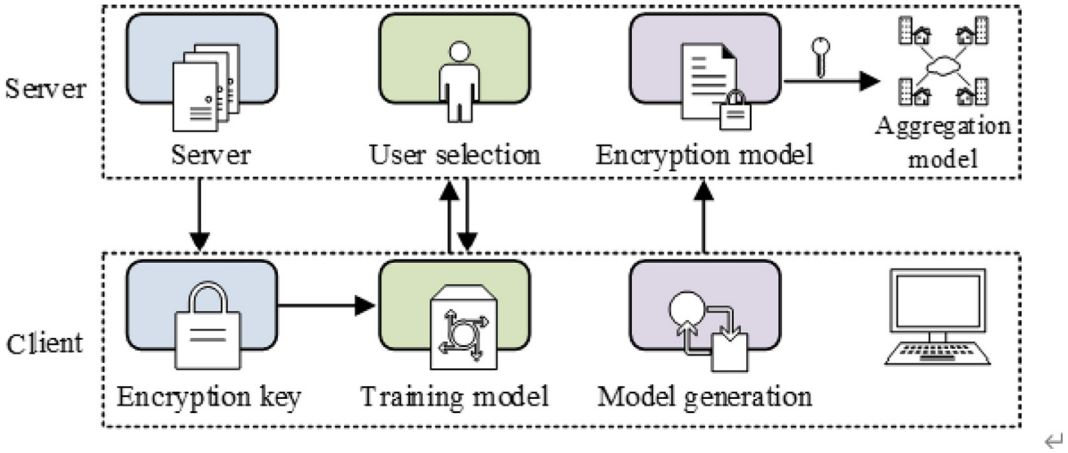

Fig. 3. Decentralized key security management pattern.

consensus mechanism to realize the decentralization of FL data aggregation, so that all participants can track the training progress in real time. In addition, the pattern parameters are encrypted in combination with data encryption to prevent the trouble that the plaintext pattern may cause privacy disclosure during the collaborative training process, and ensure that the BC subject cannot gain the pattern information of other subscribers. The data security aggregation pattern based on formula encryption and block chain (FE-BC) is displayed in Fig. 4. Subscribers use personal keys to encrypt data x_i and transmit ciphertext $Enc(x_i)$ to the BC for data sharing, while ciphertext information can only be gained using decryption keys. After the introduction of BC, the traditional federal learning and training has been changed into a cooperative training mode, and each participant has become a computing node in the BC network. Before the collaborative training starts, the initial node will first create a Genesis block, write the initial pattern trained by the public training dataset into the Genesis block, and broadcast it to the entire network node. The local equipment uploads the training data to the computing node, and write it to the candidate block after verification. Each computing node will win the qualification of new blocks through election, and then broadcast candidate blocks in the network. When the candidate blocks are recognized by other computing nodes, they will be created as new blocks. The subscriber transmits the trained and encrypted local pattern to the accounting node of the federation chain and generates a new block. The federation chain will synchronize the new block to each computing node through the point-to-point network. When the local pattern block meets the conditions of the smart contract, the endorsement node of the alliance chain will aggregate the patterns. After reaching a consensus, the polymerized global pattern will be sent to the sorting node for updating, and the quantity of iterations of the network will be updated, so that the local pattern under the original number of iterations can no longer be uploaded.

4. Edge asynchronous efficient communication framework design

4.1. Frame structure design with edge computing

In the training process, the communication overhead between customers under the classic FL framework and between customers and cloud centrum servers is large, resulting in a remarkable reduction in the overall communication productivity of the FL framework [24]. To ameliorate the communication productivity of the FL framework, edge computing and asynchronous communication mode are introduced on the cornerstone of the classic FL communication structure, and border servers are added between the original cloud centrum server and the customer, so that the customers participating in the training do not need to directly connect with the cloud centrum server, but communicate with border servers that are close to each other according to the principle of proximity.In the communication structure, the communication between the client and the edge server is synchronous. Considering the large communication cost and delay between the edge server and the cloud center server, the communication mode between the edge server and the cloud center server is changed to asynchronous communication, hoping to improve the communication efficiency of the communication structure. The edge asynchronous communication framework is displayed in Fig. 5. The customer uses synchronous random gradient descent to select a data sample with batch size of B from the local dataset for training. The average gradient calculation formula is displayed in Eq. (6).

$$w_t = \frac{1}{B}\sum_{i=1}^{B}\nabla w_{t,i} \tag{6}$$

In Eq. (6), $w_{t,i}$ is the local gradient parameter. There is H client in total, and the border server will aggregate the gradient parameters transmitted by the customer and determine the average value of the gradient parameters. The average value calculation formula of the gradient parameters is displayed in Eq. (7).

$$w = \frac{1}{H}\sum_{i=1}^{H}\nabla w_t^i \tag{7}$$

In Eq. (7), w_t^i is the customer transmittal gradient parameter. The border server sends the polymerized gradient parameters to the customer participating in the training, and the customer updates the local pattern parameters. The update formula is displayed in Eq. (8).

$$x_{t+1} = x_t - \eta_t w \tag{8}$$

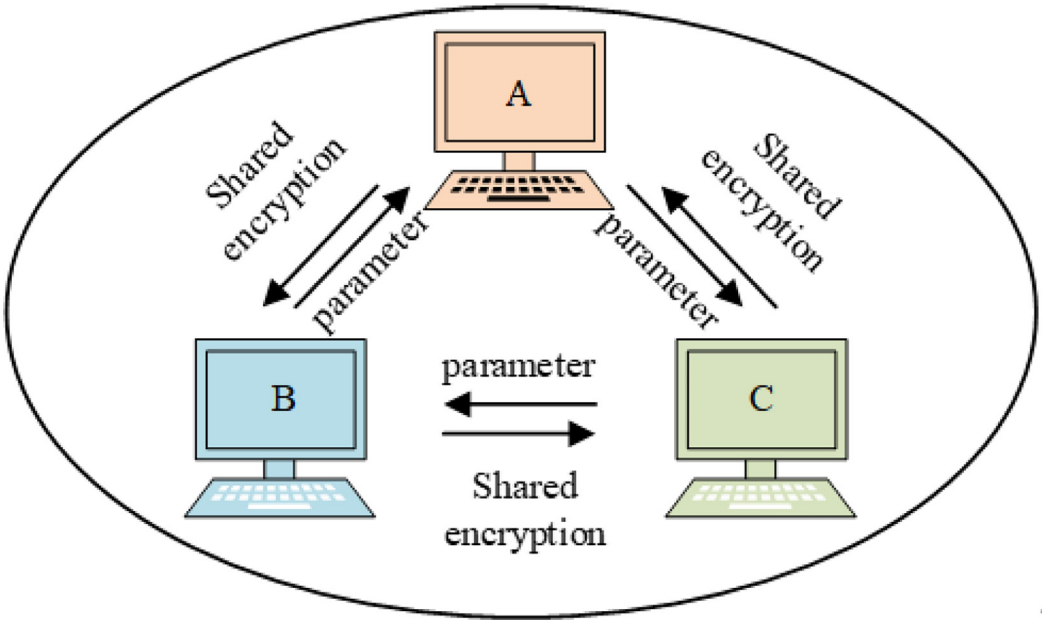

Fig. 4. Data security aggregation pattern based on BC.

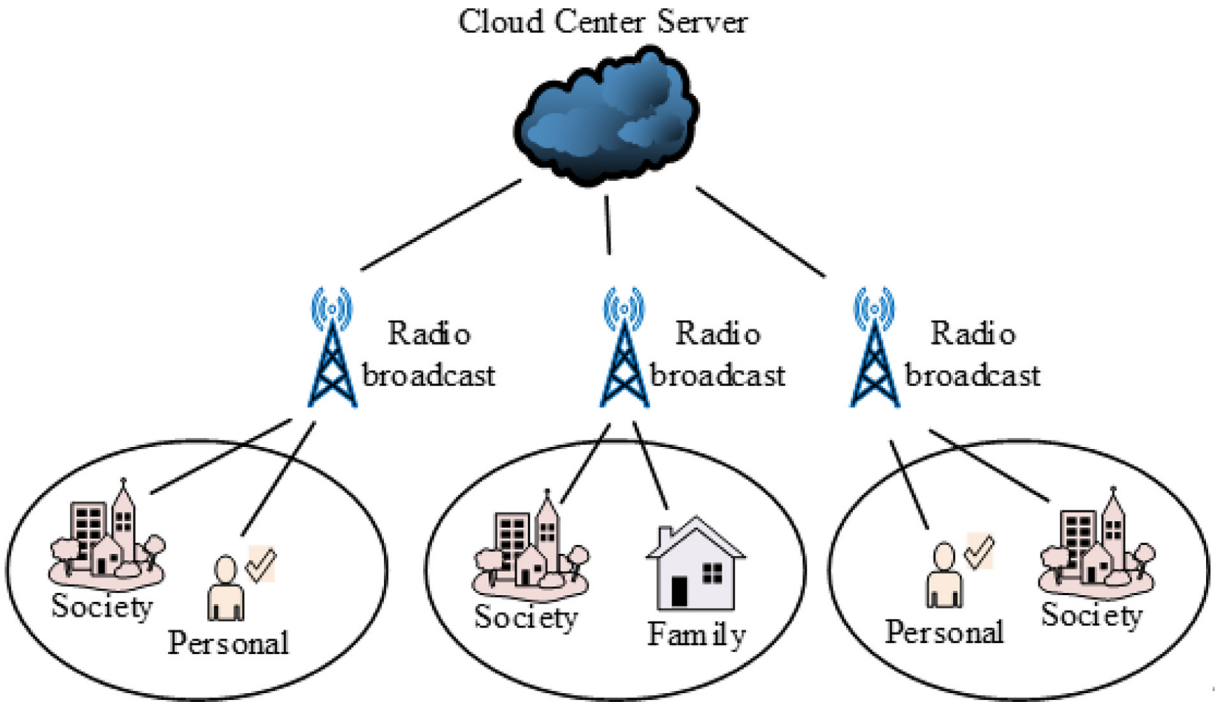

Fig. 5. Edge asynchronous communication framework.

In Eq. (8), η_t is the learning rate of the t round of training, x_{t+1} is the updated pattern parameters, and x_t is the last round of pattern parameters. The information transmittal between the customer and the border server is realized by synchronous communication. The border server will set a threshold value for the quantity of customers that upload parameters. If the quantity of customers partakeing in training and parameter upload exceeds the threshold value, the border server will not receive parameter transmittal from other customers and directly conduct aggregation operations to avoid gaining gradient parameters with gradient delay transmitted by other customers, It affects the outcomes of aggregation operations. After the aggregation operation is completed, the border server will broadcast the polymerized parameters to all customers, informing them that the iterative training has ended.

Setting the native training set of Hcustomers participating in the training as D_h . The loss outcome gained by using the pattern parameters to predict the sample data is expressed as $f_i(w) = L(x_i, y_i, w)$. Then the loss calculation formula is displayed in Eq. (9).

$$\min_{w \in R^d} f(w) = \frac{1}{n} \sum_{i=1}^{n} f_i(w) \tag{9}$$

In Eq. (9), s is the data volume of the training data set, and $w \in R^d$ is the d pattern parameters. Let P_h be the data point index set of the h customer, and the quantity of data points of the h customer is $n_h = |P_h|$. The loss formula is displayed in Eq. (10).

$$f(w) = \sum_{h=1}^{H} \frac{n_h}{n} \cdot \frac{1}{n_h} \sum_{i \in P_h} f_i(w) \tag{10}$$

The cloud centrum server sends the initialized pattern parameters to the border server, and the border server transmits them to the customer. The customer receiving the initialization parameters can participate in the training. The learning objective formula of the customer is displayed in Eq. (11).

$$\arg\min L(X, Y, \theta) = \sum_{h} a_h L_h(X, Y, \theta) \tag{11}$$

In Eq. (11), a_h is the weight of the gradient parameter gained by the customer in the aggregation operation.

4.2. ASGD based asynchronous communication

Cloud centrum servers and border servers often have large communication overhead and delay in the communication process, and different border servers have large differences in the quantity of customers and computing resources, which makes the iteration time of different border servers in the training process different [25]. If the cloud centrum server and the border server use synchronous communication, although the trouble of gradient parameter delay is solved, the communication and computing productivity of the cloud centrum server will be limited by the border server with the longest drill time. Before the border server with the longest drill time finishes uploading the parameters, the cloud centrum server will always be in a state of waiting for the parameters, resulting in a large amount of waste of computing resources. When the cloud centrum server gains all the parameters uploaded by the border server, it will update the global pattern parameters and send the updated global pattern to the border server. However, when synchronous communication is used, there is a delay gradient between the cloud centrum server and the border server, and the update of the global pattern parameters of the cloud centrum server may be affected by the upload parameters with a degree of delay [26]. Therefore, the asynchronous communication mode is further introduced into the communication framework with the border server, and the communication mode between the cloud centrum server and the border server is changed to the asynchronous communication mode to ameliorate the convergence and communication productivity of the pattern. In addition, gradient delay compensation is introduced to ensure that the gradient delay is controlled within a small range by means of delay compensation, and set a limit for gradient delay to address the trouble of delay degree that may be caused by asynchronous communication. In the asynchronous communication mode, after receiving the

gradient parameters transmitted by any border server, the cloud centrum server can directly update the global pattern without waiting for the gradient parameters of all border servers to avoid the cloud centrum server being empty. The research combines ASGD (Asynchronous Stochastic Gradient Descent) and gradient delay compensation. On the cornerstone of traditional ASGD, local gradient delay is compensated, instead of directly introducing local gradient into the global pattern, to further ameliorate the exactitude of the pattern. The gradient function is used to compensate the local gradient delay, the Taylor expansion of gradient formula $g(w_{t+\tau})$ is displayed in Eq. (12).

$$g(w_{t+\tau}) = g(w_t) + \nabla g_{(w_t)}(w_{t+\tau} - w_t) + O\left((w_{t+\tau} - w_t)^2\right) I_n \tag{12}$$

In Eq. (12), ∇g is the matrix with element g_{ij} , and$i \in [n], j \in [n]$, I_n is a n -dimensional vector, in which all elements are 1. $(w_{t+\tau} - w_t)^2$ As displayed in Eq. (13).

$$(w_{t+\tau} - w_t)^2 = (w_{t+\tau,1} - w_{t,1})^{\alpha 1} \cdots (w_{t+\tau,n} - w_{t,n})^{\alpha n} \tag{13}$$

The asynchronous random gradient descent with gradient delay compensation is displayed in Eq. (14).

$$w_{t+\tau+1} = w_{t+\tau} - \eta(g(w_t)) + \eta * \varphi g(w_t) * g(w_t) * (w_{t+\tau} - w_t) \tag{14}$$

In Eq. (14), $*$ is the element by element product, $g(w_t) * g(w_t) * (w_{t+\tau} - w_t)$ is the delay compensation gradient, and $\varphi$is the variance control parameter.

5. Effect analysis of data encryption and communication efficiency improvement methods

5.1. Data security aggregation effect verification

To evaluate the effectiveness of the research method, the feasibility of the model was tested and analyzed through experiments. TensorFlow was used for system development. The server configuration included Linux Centos 7 operating system and Hygon-C86-7159 processor. In the experiment, first analyze the accuracy of FE-BDP training, set the noise parameter sigma to 0.5, set the privacy budget eps to 8, and set the number of experimental users to 200, 800, and 1600 respectively. The user data are all from the MINST data set. The research compares the FE-BDP model with the function encryption model without differential privacy to analyze the accuracy of the model with or without differential privacy. The comparison results are shown in Fig. 6.

In Fig. 6, the pattern exactitude is compared for 200, 800 and 1600 subscribers. Fig. 6 (a) displays the exactitude change when the quantity of subscribers is 200. When the budget consumption percentage reaches 20%, the exactitude of FE-BDP rises to more than 90%. However, the exactitude of the pattern without differential privacy only rose to 80% after the budget consumption percentage reached 60%. Fig. 6 (b) and Fig. 6 (c) display the difference of pattern exactitude when the quantity of subscribers is 800 and 1600 respectively. The exactitude of FE-BDP is still significantly higher than that of the pattern without differential privacy. And in Fig. 6 (b) and Fig. 6 (c), the exactitude of FE-BDP can rise to more than 90%, while the exactitude of the pattern without differential privacy is always below 80%. However, it can be found

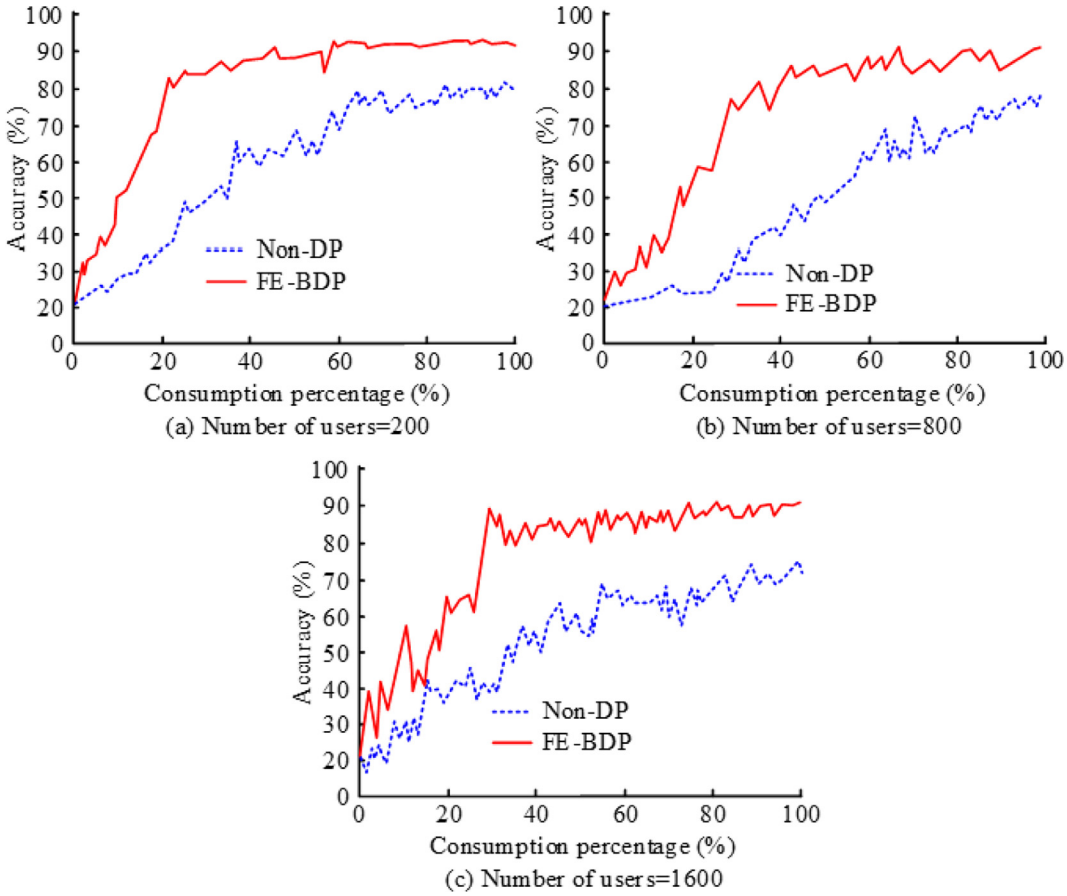

Fig. 6. Analysis of exactitude difference with and without differential privacy.

that with the build-up of the quantity of subscribers, the budget consumption percentage of FE-BDP to achieve the maximum exactitude is growing. To further evaluate the accuracy of FE-BDP, the accuracy differences between different differential privacy models are studied and compared. The FE-BDP is compared with the classic FL-DP model and the location-based service query of differential privacy (DP-LQ) model. The evaluation results are shown in Fig. 7.

Fig. 7 (a) displays the comparison outcomes of exactitude when the quantity of subscribers is 200. The exactitude of FE-BDP is significantly higher than that of other patterns, and its exactitude rate reaches more than 95%. Although the exactitude rate of the other two patterns can reach more than 95%, it is still lower than that of FE-BDP. Fig. 7 (b) and Fig. 7(c) display the comparison outcomes of pattern exactitude when the quantity of subscribers is 800 and 1600. The exactitude of FE-BDP is always higher than the other two patterns based on different number of subscribers. With the increasing number of subscribers, the exactitude of the pattern is decreasing. The above outcomes display that FE-BDP has a high exactitude rate. Although it will decrease with the build-up of the quantity of subscribers, it can still keep the exactitude rate above 90% for a long time.

FE-LDP is proposed in the study to avoid collusion and theft of user privacy information. In order to evaluate the performance of this method, loss value analysis is used to evaluate and compare FE-LDP with classic FL-DP model, DP-LQ model and FE-LDP model. The results are shown in Fig. 8.

Fig. 8 compares the difference of loss value changes between FE-LDP and other differential privacy. With the increasing number of iterations, the loss value of FE-LDP decreases ceaselessly, and finally decreases to below 0.05. It can be seen from the change of loss value of each pattern that the classical FL-DP loss value displays a growing trend, which means that under the current experimental environment, the performance of the traditional FL-DP pattern is poor. From the above outcomes, it can be found that compared with other patterns, FE-LDP has a smaller loss value, which means that FE-LDP has a smaller loss in subscriber privacy information encryption and can effectively implement information encryption.

A BC based aggregation technique is advanced to avoid the intrusion attack caused by a single point of server failure. At the

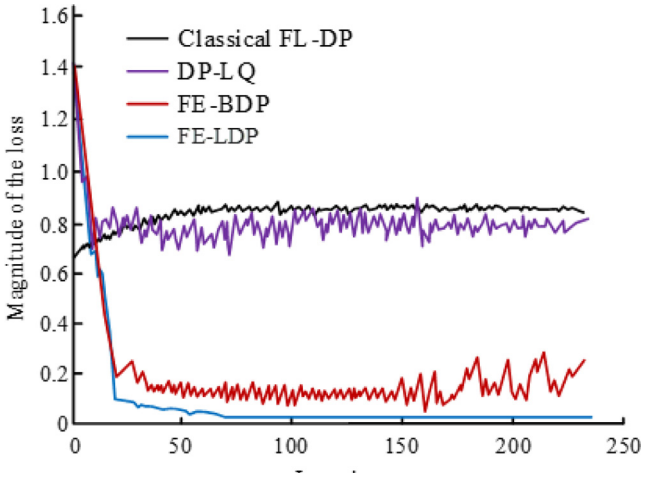

Fig. 8. Loss value evaluation of FE-LDP pattern.

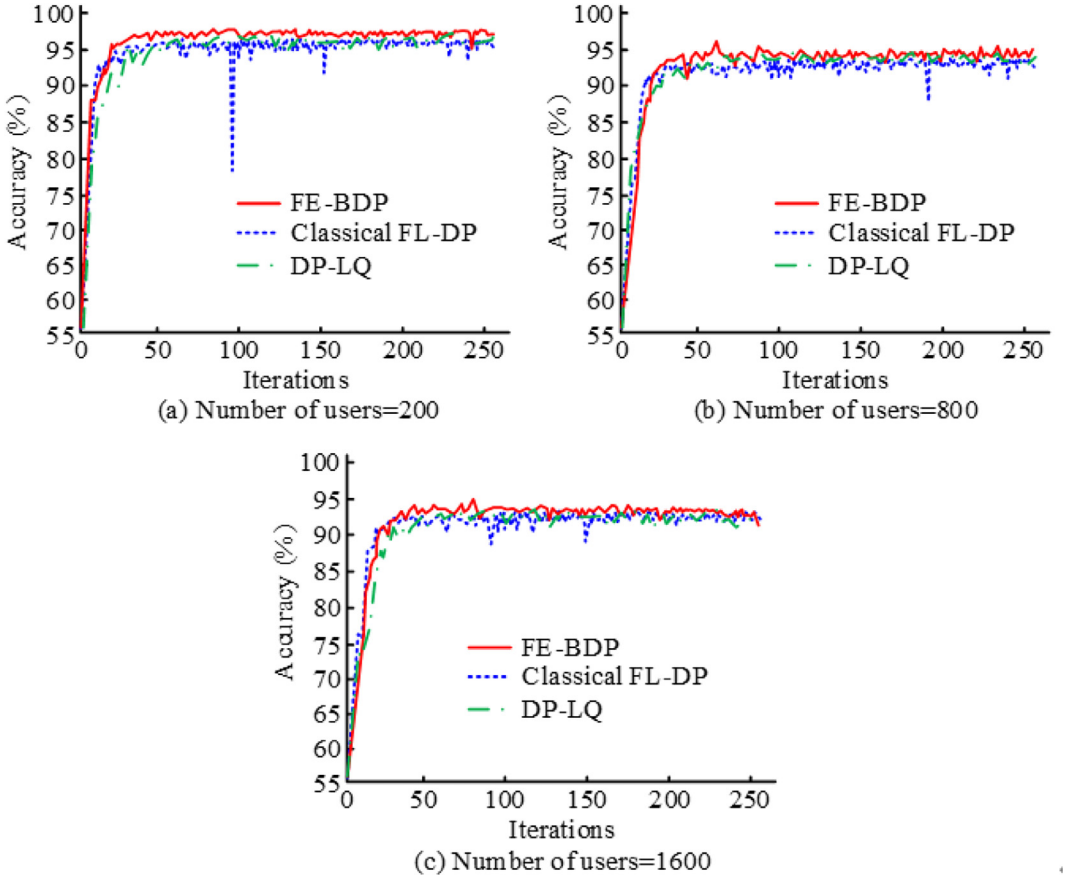

Fig. 7. FE-BDP exactitude evaluation.

same time, to avoid the leak of subscriber privacy in collaboration, a FE-BC encryption pattern based on formula encryption is advanced. To analyze the application performance of FE-BC, the aggregation time of FE-BC in ceaseless iteration is studied and evaluated, as displayed in Fig. 9.

Fig. 9 displays the difference between the pattern aggregation time and training under different subscriber numbers. The pattern aggregation time is always kept below 1.0s in the ceaseless iterative calculation process, indicating that the pattern aggregation cost is small. Comparing the training practice of the pattern, it can be found that the drill time of the pattern is more than 1.0s, which is higher than the aggregation time of the pattern, indicating that the aggregation ability of the pattern is better. Therefore, with the increasing number of subscribers, the aggregation time of FE-BC does not change significantly, and the overall operation is stable. The above outcomes display that FE-BC has a fast polymerization time and is relatively stable during operation.

5.2. Communication productivity improvement verification

An edge asynchronous efficient communication framework is advanced to ameliorate the data communication processing capability. Therefore, to verify its communication performance, a communication productivity evaluation experiment is advanced. In the experiment, the experimental environment is first built, as displayed in Table 1. Use the experimental environment in Table 1 to carry out the communication framework, compare the convergence between different communication frameworks, and evaluate the information processing capacity of each communication framework, as displayed in Fig. 10.

In Fig. 10, the convergence between heterogeneous patterns, edge patterns and edge asynchronous patterns is compared. Fig. 10 (a) displays the MSE change of heterogeneous patterns. It can be found that with the increasing number of iterations, the MSE value of heterogeneous patterns displays a decreasing trend, and gradually reduces to the minimum after 500 iterations. Fig. 10 (b) is the edge pattern. The MSE value of the edge pattern also displays a decreasing trend with the increasing number of iterations. Fig. 10 (c) displays the MSE value change of the edge asynchronous pattern. It can be found that with the increasing number of iterations, the MSE value of the edge asynchronous pattern decreases ceaselessly. The above outcomes display that the edge asynchronous efficient communication framework advanced in the study has good convergence, and it can achieve fast stability and avoid the increase of data error caused by communication fluctuations. Finally, in order to evaluate the communication efficiency of the edge-asynchronous efficient communication framework in practical application, the research took a city mall, parking lot, university and hospital as the research site, and carried out the actual test of the edge-asynchronous efficient communication framework, and compared with the decentralized framework and asynchronous gradient reduction framework. The analysis results are shown in Table 2.

In Table 2, in different scenarios, the exactitude of the edge asynchronous efficient communication framework advanced in the study is more than 90%, and the calculation loss rate is small. The maximum value appears in the hospital environment, which is 0.68%. The reason is that the communication loss rate of the hospital is higher than that of other scenarios due to the complex

Table 1
Software and hardware information.

Experimental environment	Edition
Operating system	Windows 10
Processor	Inter Core i5, 6200CPU, 2.4 GHz
ECS	Centos7.3
Operating language	Python 3.6

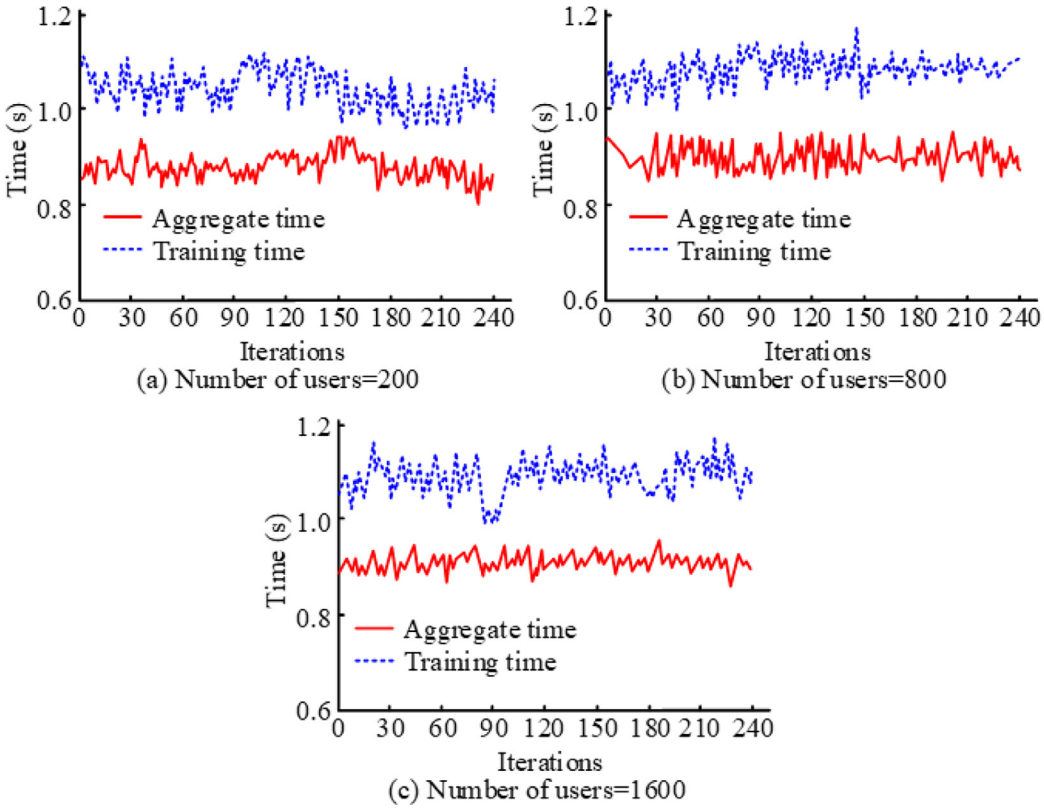

Fig. 9. FE-BC time cost evaluation.

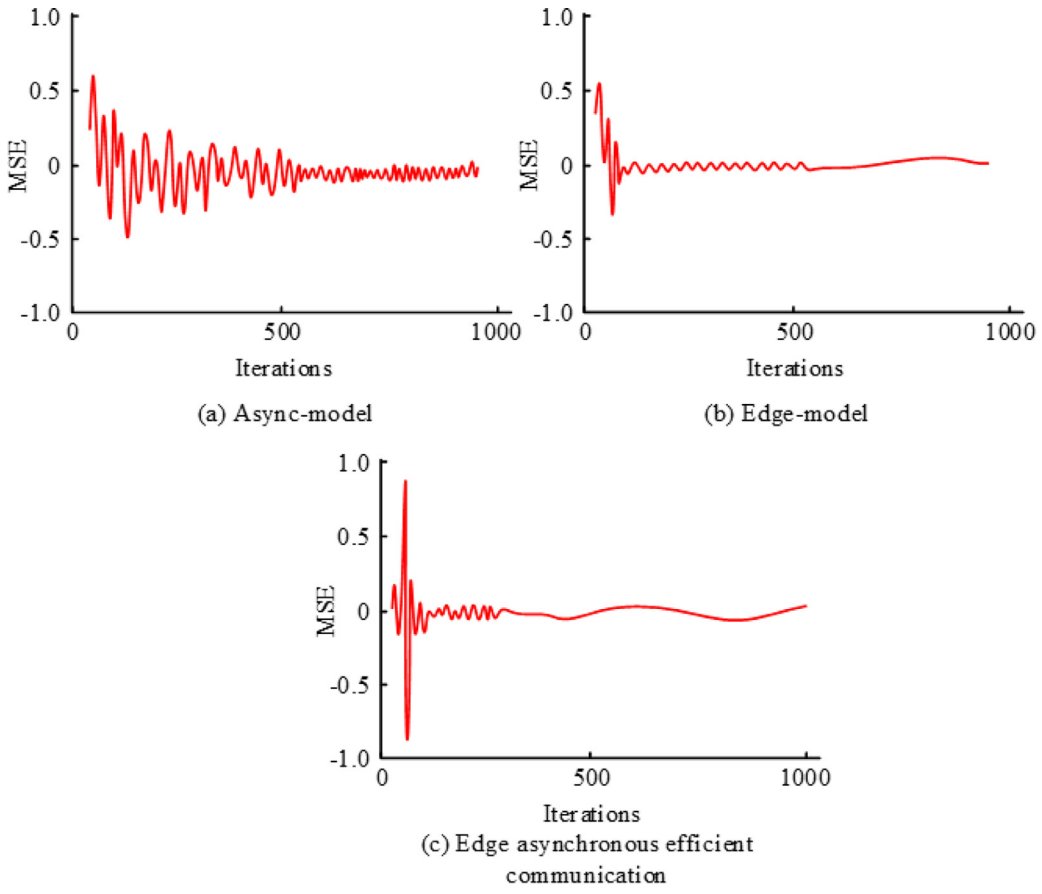

Fig. 10. Convergence Analysis of Edge Asynchronous Efficient Communication Framework.

Table 2
Comparative test of communication framework application.

Frame		Market	Parking lot	Colleges and universities	Hospital
Edge - asynchronous framework	Accuracy (%)	0.9274	0.9387	0.9365	0.9291
	Loss rate (%)	0.0043	0.0066	0.0057	0.0068
	Communication cost (s)	1342.07	1315.29	1318.66	1331.11
Decentralized framework	Accuracy (%)	0.8567	0.8439	0.8531	0.8572
	Loss rate (%)	0.1433	0.1561	0.1469	0.1428
	Communication cost (s)	1823.24	1803.42	1816.79	1827.51
Asynchronous gradient descent frame	Accuracy (%)	0.8731	0.8716	0.8843	0.8715
	Loss rate (%)	0.1269	0.1284	0.1157	0.1285
	Communication cost (s)	1713.05	1692.37	1701.38	1710.25

environment of the hospital. Analyzing the communication overhead of the edge asynchronous efficient communication framework, we can find that in different scenarios, the time overhead is less than 1350s, and we can know that the communication overhead of the communication framework in different scenarios is not significantly different. The above outcomes display that the edge asynchronous efficient communication framework advanced in the study has a high exactitude and low loss rate, and can deal with different urban scenarios, while displaying consistent communication overhead, which further displays that the edge asynchronous efficient communication framework has a high stability.

6. Conclusion

Smart city is the fundamental development direction of the current city, so it is crucial to ensure the data security in the development of smart city. To ameliorate the preservation of smart city data, a data encryption technique under the FL is advanced. This technique uses formulaal encryption and differential privacy preservation technology to avoid customer data leak. At the same time, to ameliorate the development rate of smart city, an edge asynchronous framework is advanced to ameliorate communication productivity. The pattern verification displays that the advanced FE-BDP encryption pattern has significantly ameliorated its exactitude after the introduction of differential privacy technology, and can maintain the exactitude of more than 90% for a long time. Compared with other encryption patterns, FE-BDP has a higher exactitude, which is more than 95%. In addition, the loss value of the FE-LDP pattern advanced in the study is significantly smaller than that of other patterns, and can be stabilized to below 0.05. At the same time, the aggregation time of the BC based data aggregation technique is less than 1.0s. Finally, evaluate the communication productivity improvement effect of the edge asynchronous framework. The outcomes display that the edge asynchronous framework has significantly better convergence,

and its communication overhead can be reduced to 1315.29s at the lowest. The above outcomes display that under the FL, the introduction of formulaal encryption and differential privacy preservation technology can effectively achieve data privacy preservation, and can ameliorate communication productivity with the edge asynchronous framework. However, the study did not consider the city data that is growing with time. Therefore, in the subsequent study, we will advance a constantly updated data encryption and communication productivity improvement technique based on the change of city data dimension.

CRediT authorship contribution statement

Zhen Kuang: Conceptualization, Methodology, Software, Visualization, Formal analysis, Data curation, Writing – original draft, Investigation. **Chaoyang Chen:** Supervision, Software, Validation, Writing – review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Key Rand&D Program of China for International S&T Cooperation Projects (2019YFE0118700), National Natural Science Foundation of China (62222306, 61973110), Hunan Young Talents Science and Technology Innovation Project (2020RC3048), Natural Science Found for Distinguished Young Scholars of Hunan Province (2021JJ10030), the Foundation of State Key Laboratory of Digital Manufacturing Equipment and Technology DMETKF2022023.

References

[1] Nica E. Urban big data analytics and sustainable governance networks in integrated smart city planning and management. Geopolitics History Int Rel 2021;13(2):93–106.
[2] Yang Z, Chen M, Saad W, et al. Energy efficient federated learning over wireless communication networks[j]. IEEE Trans Wireless Commun 2020;20 (3):1935–49.
[3] Dinh CT, Tran NH, Nguyen MNH, et al. Federated learning over wireless networks: Convergence analysis and resource allocation. IEEE/ACM Trans Networking 2020;29(1):398–409.
[4] Yu S, Chen X, Zhou Z, et al. When deep reinforcement learning meets federated learning: Intelligent multitimedimension resource management for multiaccess edge computing in 5g ultradense network. IEEE Internet Things J 2020;8(4):2238–51.
[5] Chen M, Yang Z, Saad W, et al. A joint learning and communications framework for federated learning over wireless networks. IEEE Trans Wireless Commun 2020;20(1):269–83.
[6] Saračević M, Adamović S, Maček N, et al. Cryptographic keys exchange pattern for smart city applications. IET Intel Transport Syst 2020;14(11):1456–64.
[7] Xi P, Zhang X, Wang L, et al. A review of block chain-based secure sharing of healthcare data. Appl Sci 2020;12(15):7912–23.
[8] Yin L, Feng J, Xun H, et al. A privacy-preserving federated learning for multiparty data sharing in social iots. IEEE Trans Network Sci Eng 2021;8 (3):2706–18.
[9] Ou W, Zeng J, Guo Z, et al. A homomorphic-encryption-based vertical federated learning scheme for rick management. Comput Sci Inf Syst 17(3), 1736–1746, 819–834.
[10] Arachchige PCM, Bertok P, Khalil I, et al. A trustworthy privacy preserving framework for machine learning in industrial iot systems. IEEE Trans Industr Inf 2020;16(9):6092–102.
[11] Jia B, Zhang X, Liu J, et al. Block chain-enabled federated learning data preservation aggregation scheme with differential privacy and homomorphic encryption in iiot. IEEE Trans Industr Inf 2021;18(6):4049–58.
[12] Yu R, Li P. Toward resource-efficient federated learning in mobile edge computing. IEEE Network 2021;35(1):148–55.
[13] Sun H, Li S, Yu FR, et al. Toward communication-efficient federated learning in the internet of things with edge computing. IEEE Internet Things J 2020;7 (11):11053–67.
[14] Duan M, Liu D, Chen X, et al. Self-balancing federated learning with global imbalanced data in mobile systems. IEEE Trans Parallel Distrib Syst 2020;32 (1):59–71.
[15] Zhong R, Liu X, Liu Y, et al. Mobile reconfigurable intelligent surfaces for noma networks: Federated learning approaches. IEEE Trans Wireless Commun 2022;21(11):10020–34.
[16] Kaur A, Kumar K. Energy-efficient resource allocation in cognitive radio networks under cooperative multi-agent pattern-free reinforcement learning schemes. IEEE Trans Netw Serv Manage 2020;17(3):1337–48.
[17] Nguyen DC, Ding M, Pham QV, et al. Federated learning meets block chain in edge computing: Opportunities and challenges. IEEE Internet Things J 2021;8 (16):12806–25.
[18] Le THT, Tran NH, Tun YK, et al. An incentive mechanism for federated learning in wireless cellular networks: An auction approach. IEEE Trans Wireless Commun 2021;20(8):4874–87.
[19] Lim WYB, Huang J, Xiong Z, et al. Towards federated learning in uav-enabled internet of vehicles: A multi-dimensional contract-matching approach. IEEE Trans Intell Transp Syst 2021;22(8):5140–54.
[20] Lu Y, Huang X, Dai Y, et al. Differentially private asynchronous federated learning for mobile edge computing in urban informatics. IEEE Trans Industr Inf 2019;16(3):2134–43.
[21] Khan LU, Pandey SR, Tran NH, et al. Federated learning for edge networks: resource optimization and incentive mechanism. IEEE Commun Mag 2021;58 (10):88–93.
[22] Chen D, Hong CS, Wang L, et al. Matching-theory-based low-latency scheme for multitask federated learning in mec networks. IEEE Internet Things J 2021;8(14):11415–26.
[23] Samarakoon S, Bennis M, Saad W, et al. Distributed federated learning for ultra-reliable low-latency vehicular communications. IEEE Trans Commun 2019;68(2):1146–59.
[24] Jasim NA, TH H, Rikabi SAL. Design and implementation of smart city applications based on the internet of things. Int J Interactive Mobile Technol 2021;15(13):4–15.
[25] Lu Y, Huang X, Zhang K, et al. Low-latency federated learning and block chain for edge association in digital twin empowered 6g networks. IEEE Trans Industr Inf 2020;17(7):5098–107.
[26] Lim WYB, Luong NC, Hoang DT, et al. Federated learning in mobile edge networks: A comprehensive survey. IEEE Commun Surveys Tutor 2020;22 (3):2031–63.