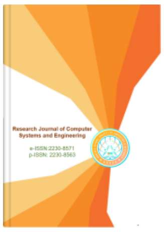
Research Journal of Computer Systems and Engineering

ISSN: 2230-8571, 2230-8563

Volume 03 Issue 01 - 2022 (January to June)

Page 21:28



Blockchain Based 5g Heterogeneous Networks Using Privacy Federated Learning with Internet of Things

Dr.Nikhitha Yathiraju
Computer Science & Communication Engineering
University of Cumberlands
ORCHID ID-0000-0003-1908-3267
nikhitha.yathiraju6743@ucumberlands.edu

Article History

Received: 22 January 2022
Revised: 14 April 2022
Accepted: 19 May 2022

Abstract

Federated learning, which is a distributed machine learning approach for preserving privacy and is thus widely used in numerous privacy issue applications, is involved in the area where privacy is of higher importance. To protect the privacy of users' local gradients while conducting federated learning, elliptical curve cryptography with block chain-based federated learning (ECC-BFL) is proposed here.Considerable consideration is given to factors including categorization accuracy, running time, communication overhead, computation overhead, and transaction speed. The values for these parameters are compared to three established techniques: the Biparing Method (BM), the Homomorphic Cryptosystem (HC), and the Multiple Authorities with Attribute-Based Signature scheme (MA-ABS). Additionally, a proposed Elliptical Curve Cryptography with Blockchain-based Federated Learning (ECC-BFL)technique is also considered.The suggested ECC-BFL was able to accomplish 95% classification accuracy, 65 sec. of operating time, 76% communication overhead, 63% calculation overhead, and 92% transaction speed.

Keywords: blockchain, 5G network, federated machine, privacy preservation, registration

1. Introduction

In machine learning (ML) approaches used traditionally, the model's accuracy and efficiency are based on the data trained and computing power of the centralized server. Precisely, with traditional ML approaches, central sever is the storage for user data which is used for both training and testing where wide range of ML models are developed in due course [1]. Federated Learning (FL) introduced in [2] has emerged with a solution by addressing this issue. FL provides privacy for user data where data are decentralized from the central server to end-devices [3]. Privacy preservation offers possibilities to influence the benefits of AI achieved by efficient machine learning models across various domains [4]. Besides privacy, FL permits ML to be used in smaller domains where only inadequate training data is available to construct a standalone ML model.

2. Related works

Privacy in IoT based systems can be preserved by using the method of anonymization and thus numerous researchers have employed this technique to preserve privacy in blockchain-based IoT applications. In [5], an App for smart mobiles named Healthcare Data Gateway (HDG) was developed based on blockchain with MPC approach.In [6], and Attribute-Based Signature method with Multiple Authorities (MA-ABS) was introduced for the blockchain-based healthcare system. In [7], a cross-domain medical image sharing method was designed where patients were able to access their medical data electronically in a secured way.In [8], a simplified Indicator Centric Schema (ICS) was developed which could easily manage any type of healthcare data by using a single simple "table". In [9, 10], the bilinear pairing concept was involved to secure the location details and identity of the patient.

3. System model

Numerous researches are carried and several operational challenges are experienced with growing consumer-related techniques. Architecture of the proposed ECC-BFL healthcare systems is depicted in Figure 1.

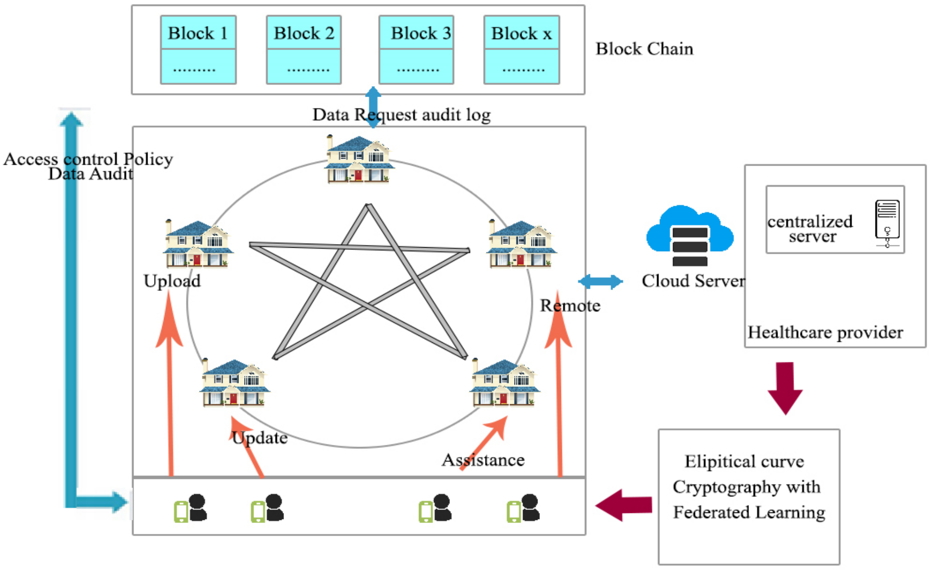


Figure-1 Architecture of ECC-BFL

3.1 Construction of blockchain network

Blockchain, sequential blocks, includes a list of valid complete transaction records. Blocks are interconnected with one another using a hash value (reference), and thus a chain is formed. First block is genesis block and one which is before given block is the parent of the given block.

- Block version specifies the validation rules;
- Previous block hash provides the hash value of that block
- Timestamp describes time taken to create current block

- The nonce is a random field with 4-bytes which is adjusted while mining for calculating each hash thereby providing a solution to the PoW puzzle

- Target hash is the threshold hash value of a new valid block that identifies difficulties in the PoW puzzle

3.2 Elliptical Curve cryptography with federated Learning

- The global model parameters are initialized on the server which is then transferred to every client connected to the network.

- These clients then learn the transferred model on its data for various training epochs. When this process is completed, updated model parameters or gradients are forwarded to the server. Gradients are the difference between the downloaded and updated models. It is to be noted that the scale of training data may differ and computational resources may be unbalanced for clients. Thus, the server fails in receiving the information uploaded by clients.

- The received uploads are aggregated by the server which can be either in synchronous or asynchronous mode with which global method is updated.

- Until converging, above two processes are repeated.

Generally, a neural network is defined as a function f(x, ω) = y', where x, y'represents inputs of the user and its respective outputs respectively through function f with ω as its parameter. With no generality loss, let every data record be (x,y), an observation pair, and D = {⟨xi, yi⟩, i = 1, 2, · · · T} represents complete training set. The loss function (Lf) for the training set is described as eq. (1)

$$\mathrm{Lf}\,(\mathrm{D}, \omega) = \frac{1}{D}\sum_{(xi,yi)\epsilon D} Lf\,(\text{ xi, yi, } \omega)$$

$$\omega^{j+1} \leftarrow \omega^{\,j} - \lambda\nabla \mathrm{L}_{\mathrm{f}}\,(\mathrm{D}^{\,j}, \omega^{\,j}) \qquad (1)$$

hereω ^{j}, D and λ represent the parameters after iteration j, an arbitrary subset of D, and learning rate parameters respectively. In the federated learning technique used here, every user n ∈ N has a private local dataset Dn, which is trained using a specific neural network where D=$\sum_{n\in N} Dn$.N j is an arbitrary subset selected by the server at iteration j, and then every user n of N j selects subset Dj at random (n⊆Dn)for executing stochastic gradient descent. Hence, the parameter update is given byeq. (2)

$$\omega^{j+1} \leftarrow \omega^{\,j} = \lambda\,\frac{\sum_{n\in N} pj}{\sum_{n\in N} Dj} \qquad (2)$$

where ρ j n = |Dj n|∇Lf (D^{j} n, ω ^{j}) is calculated by every user which is shared with cloud server.

Generate public/secret keys by eq. (3)

δ, ρ), (*NPKn ,NSKn*), (*PPKn , PSK)* *(3)*

user*n*, (*n 2 U, jUj= N*)

Select a random number *βn*

```
Generate the shares of $\beta n$
Receive messages from at least $t$ users
    $HF(xn) = (An,Bn) = (gHF\_;\_(xn), hHF\_;\_(xn))$.
Receive messages from at least $t$ users
Broadcast the list $U3$ to each user $2$ $U2$.
    Check whether $U3$ _ $U2$ and $jU3j$ _ $t$.
Functional key assignment ()
{
    If the patient confirm transaction over blockchain then
    Generate a key kusing ECC
A← get values from ECC
'a' sends request with appendend value to b
b computes from a to b
b sends reply with c
The server check for c
If
    c=ab
return to server
else
    start transaction
}
```

4. Experimental analysis

The experimental results are implemented, and the classification accuracy, running time, communication overhead, computation overhead, and transaction speed are the parameters used for analysis. ECC-BFL is proposed, and the values obtained for these parameters are compared against three established methods: BM, HC and MA-ABS.

- Classification accuracy

Total users taking part in training and for every user,the local gradient sizeis considered. Generally, the accuracy (A)produced by the model is ratio of the Total Gradients (TG) to the Total Users (TU) participating in training by eq. (4).

$$A=\frac{TG}{TU} \qquad (4)$$

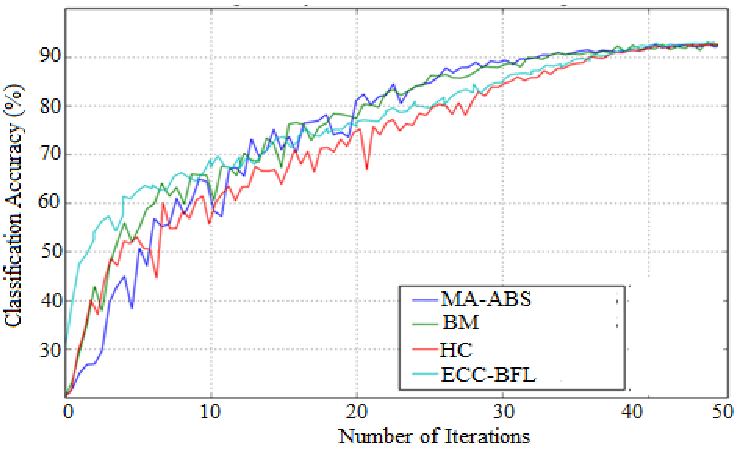


Figure-2 Analysis of classification accuracy

- Running time is the time taken by the application server to respond to the user request. Some factors which affect this time are number of users, network bandwidth, type and number of requests made by the user and mean of the thinking time.

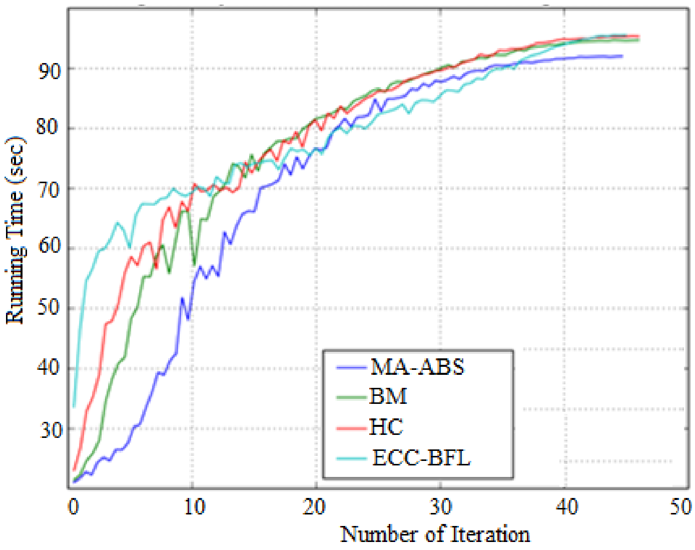


Figure-3 Analysis of running time

Figure 2 and 3illustrates the classification accuracy and running time for various iterations. When the number of iterations is increased, the system performs better and the classification accuracy of the model is also improved. It is even more clear that when the number of gradients is increased, high accuracy is produced by the model.

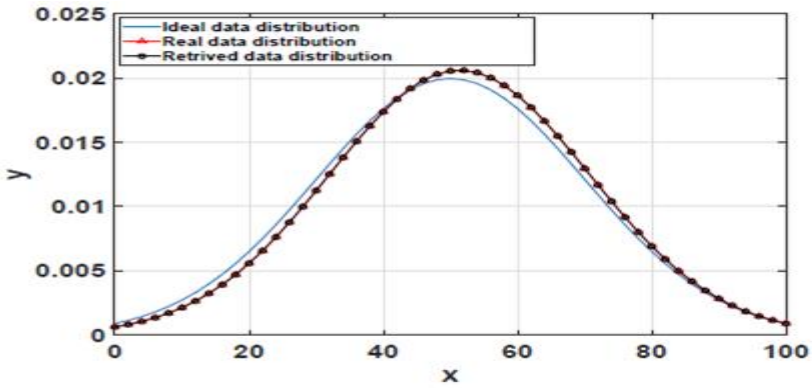


Figure-4 Analysis of data distribution

Figure 4 illustrates the data selected at random which are discrete points where the real distribution differs slightly from the ideal distribution. At the request of verification, the distribution of data uploaded which assists in generating the aggregated outputs must be identical to the real data. The aggregated outputs are further used while calculating the mean and variance of the data uploaded.

- Communication overhead

Total messages and data that have to be exchanged are estimated in communication overhead. Data involved in the proposed model occupies 2 bytes and 5 bytes for identifier and time stamp respectively where q, elliptic curve point and signature occupies 20 bytes each.

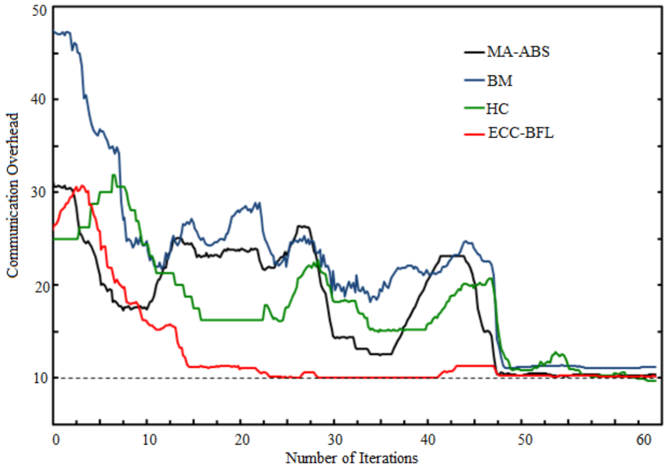


Figure-5 Analysis of Communication overhead

Figure 5 compares the communication overhead between existing MA-ABS,BM,HC and proposed ECC-BFLmethods where X axis indicates the number of iterations and Y axis the values of communication overhead. When compared, the proposed method achieves less communication overhead

- Computation overhead

The time taken by the 5G node for executing the functions involved in the authentication procedure is termed computation overhead. This reduces the congestion as well as the security risks in the core network.

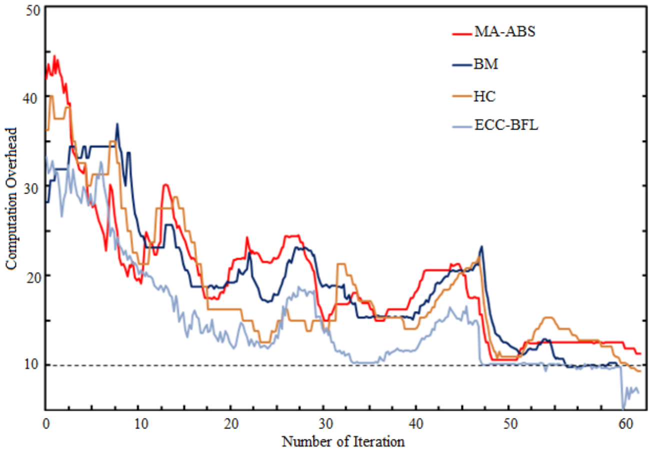


Figure-6 Analysis of Computation overhead

The figure 6 compares the computation overhead between existing MA-ABS,BM,HC and proposed ECC-BFL where X axis indicates the number of iterations and Y axis the values of computation overhead. When compared, the proposed method achieves less computation overhead. Table-1 indicates the comparison of existing BM , HC , MA-ABS and proposed method ECC-BFL

Table-1 Comparison of existing and proposed method

Parameters	BM	HC	MA-ABS	ECC-BFL
classification accuracy (%)	90	93	91	95
running time (sec)	80	72	73	65
Communication overhead (%)	85	80	81	76
Computation overhead (%)	72	68	69	63
transaction speed (%)	85	90	88	92

5. Conclusion

Blockchain potentially transforms conventional healthcare industries.Deep neural networks have failed to produce apt solutions thus ECC-BFL is proposed which helps in verifying the calculation results of the server for every user. Moreover, ECC-BFL supports users to drop out of training method. Further, experiments on real-time data practically demonstrated performance of ECC-BFL approach. 95% classification accuracy, 65 seconds of running time, 76% communication overhead, 63% computation overhead, and 92% transaction speed were all accomplished by the suggested ECC-BFL technique.In future, the reduction of communication overhead has to be focused on the entire protocol.

Reference

[1] Cabaj, K., Caviglione, L., Mazurczyk, W., Wendzel, S., Woodward, A., & Zander, S. (2018). The new threats of information hiding: The road ahead. *IT professional*, *20*(3), 31-39.

[2] Xu, C., Lin, H., Wu, Y., Guo, X., & Lin, W. (2019). An SDNFV-based DDoS defense technology for smart cities. *IEEE Access*, *7*, 137856-137874.

[3] Ning, Z., Dong, P., Wang, X., Rodrigues, J. J., & Xia, F. (2019). Deep reinforcement learning for vehicular edge computing: An intelligent offloading system. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(6), 1-24.

[4] Zhang, L., Wu, Q., Domingo-Ferrer, J., Qin, B., & Hu, C. (2016). Distributed aggregate privacy-preserving authentication in VANETs. *IEEE Transactions on Intelligent Transportation Systems*, *18*(3), 516-526.

[5] Sun, G., Sun, S., Sun, J., Yu, H., Du, X., &Guizani, M. (2019). Security and privacy preservation in fog-based crowd sensing on the internet of vehicles. *Journal of Network and Computer Applications*, *134*, 89-99.

[6] Servos, D., & Osborn, S. L. (2017). Current research and open problems in attribute-based access control. *ACM Computing Surveys (CSUR)*, *49*(4), 1-45.

[7] Sultana, N., Chilamkurti, N., Peng, W., &Alhadad, R. (2019). Survey on SDN based network intrusion detection system using machine learning approaches. *Peer-to-Peer Networking and Applications*, *12*(2), 493-501.

[8] Zhao, J., Chen, Y., & Zhang, W. (2019). Differential privacy preservation in deep learning: Challenges, opportunities and solutions. *IEEE Access*, *7*, 48901-48911.

[9] Yue, X., Wang, H., Jin, D., Li, M., & Jiang, W. (2016). Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. *Journal of medical systems*, *40*(10), 1-8

[10] Yazdinejad, A., Parizi, R. M., Dehghantanha, A., &Choo, K. K. R. (2019). Blockchain-enabled authentication handover with efficient privacy protection in SDN-based 5G networks. *IEEE Transactions on Network Science and Engineering*, *8*(2), 1120-1132.