Sharded Blockchain for the Scalability and Privacy of Healthcare Data in Federated Learning ⋆

Jahanzeb Shahid[1,*,†], Stelvio Cimato[2,†]

Abstract

The application of Federated Learning (FL) techniques to electronic health records (EHR) is gaining more and more attention as a method to extract valuable information that has the potential to enhance the decision-making process within the healthcare domain. It becomes a useful tool when integrated with blockchain technology, which provides some properties such as immutability and traceability that are useful to enhance the security of such applications. In this research, we propose an architecture to address the scalability and privacy problem for sharing health data adopting a sharding-based blockchain framework. We then describe a possible implementation relying on Hyperledger fabric.

1. Introduction

Federated learning (FL) is a new type of distributed machine learning that does not consolidate data on cloud data centers during training [1]. A typical federated learning method combines centralized servers and connected devices and systems interacting over the Internet [2]. The model-first approach in FL reduces data communication costs, optimizing bandwidth and reducing latency, increases privacy preservation. Federated learning (FL) model parameter aggregation makes the entire model dependent on the central FL server. A central server failure can cause a single point of failure (SPoF) situation, potentially leading to a denial of service (DoS) attack [3]. Other FL-related security challenges include model inversion, model poisoning, inference, data label poisoning, and device authentication and authorization. In this paper we present a blockchain-based architecture for healthcare data that leverages sharding to address scalability issues related to the large amount of healthcare data produced by many interconnected devices, data that will be later used to train ML models. In literature, several approaches have been proposed combining FL with blockchain frameworks. In [4], authors introduced ScaleSFL, a blockchain-based technique for federated learning. The separation of off-chain federated learning components, which authenticate model modifications, facilitates interoperability. Hyperledger Fabric was used to develop a prototype system to demonstrate its feasibility. Their study found that sharding can improve validation and scale federated learning (FL). Skunk, a blockchain-based zero-trust security framework, was proposed for an IoT-based 5G/6G slicing network trying to address FL security and privacy concerns [5]. Sharding-based consensus was built in the blockchain network using 5G network slices, where shards conduct transactions and consensus independently.

jahan.shahid@unimi.it (J. Shahid); stelvio.cimato@unimi.it (S. Cimato)
0000-0002-9891-1317 (J. Shahid); 0000-0003-1737-6218 (S. Cimato)

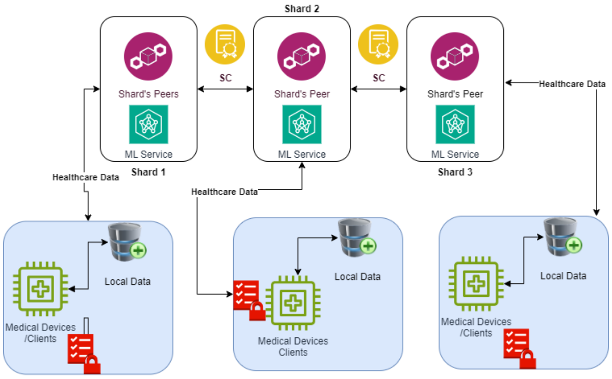

Figure 1: Sharded-Blockchain Architecture for Healthcare Data Federated Learning

2. Sharded Blockchain-Based Architecture

Sharded blockchains on the Internet of Healthcare Things (IoHT) network underpin the proposed method. The main idea is to use shards as a method to both address scalability and privacy of the produced data. When many interconnected devices produce large amount of data, as usually it occurs when multiple sensors register patients' data, shards can be used to separate and organize the collected data. In this perspective one blockchain shard can be used for each healthcare network, such as an hospital or a single department within an hospital. Each medical organization in these networks stores and processes locally the data collected by the IoHT devices, and train the model locally before sending it to the central server.

In this way storage and communication costs are reduced and throughput can be increased with the number of shards. Figure 1 shows three blockchain shards collecting healthcare data, where each peer contributes to the Federated Learning model. We are designing an implementation relying on Hyperledger fabric, developing the federated learning pipeline, creating the initial model, generating local models by each member, and creating a global model using the model averaging function.

3. Conclusion

When processing massive amounts of healthcare data on a federated learning platform, sharded blockchain and healthcare data security are intriguing study areas. The suggested technique addresses basic difficulties in centralized coordinator-based federated learning systems. We propose using FL network and blockchain-based sharding to train healthcare data. Healthcare data privacy and network scalability are important concerns. Local healthcare equipment data is used by peer nodes to generate ML models. Through blockchain, peers share locally developed models and parameters.

References

[1] J. Konecnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, D. Bacon, Federated learning: Strategies for improving communication efficiency, arXiv preprint arXiv:1610.05492 8 (2016).

[2] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečnỳ, S. Mazzocchi, B. McMahan, et al., Towards federated learning at scale: System design, Proceedings of machine learning and systems 1 (2019) 374–388.

[3] A. Qammar, A. Karim, H. Ning, J. Ding, Securing federated learning with blockchain: a systematic literature review, Artificial Intelligence Review 56 (2023) 3951–3985.

[4] E. Madill, B. Nguyen, C. K. Leung, S. Rouhani, Scalesfl: a sharding solution for blockchain-based federated learning, in: Proceedings of the Fourth ACM International Symposium on Blockchain and Secure Critical Infrastructure, 2022, pp. 95–106.

[5] E. Bandara, X. Liang, S. Shetty, R. Mukkamala, A. Rahman, N. W. Keong, Skunk—a blockchain and zero trust security enabled federated learning platform for 5g/6g network slicing, in: 2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), IEEE, 2022, pp. 109–117.