Future Generation Computer Systems 129 (2022) 380–388

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A framework for privacy-preservation of IoT healthcare data using Federated Learning and blockchain technology

Saurabh Singh [a], Shailendra Rathore [b,d], Osama Alfarraj [c], Amr Tolba [c], Byungun Yoon [a,*]

[a] Department of Industrial and System Engineering, Dongguk University, Seoul, South Korea
[b] School of Engineering, Computing and Mathematics, University of Plymouth, UK
[c] Computer Science Department, Community College, King Saud University, Riyadh 11437, Saudi Arabia
[d] Division of Cyber Security, School of Design and informatics, Abertay University, UK

ARTICLE INFO

Article history:
Received 14 March 2021
Received in revised form 8 November 2021
Accepted 26 November 2021
Available online 12 December 2021

Keywords:
Federated Learning
Privacy-preserving
Blockchain
Internet-of-Things

ABSTRACT

With the dramatically increasing deployment of IoT (Internet-of-Things) and communication, data has always been a major priority to achieve intelligent healthcare in a smart city. For the modern environment, valuable assets are user IoT data. The privacy policy is even the biggest necessity to secure user's data in a deep-rooted fundamental infrastructure of network and advanced applications, including smart healthcare. Federated learning acts as a special machine learning technique for privacy-preserving and offers to contextualize data in a smart city. This article proposes Blockchain and Federated Learning-enabled Secure Architecture for Privacy-Preserving in Smart Healthcare, where Blockchain-based IoT cloud platforms are used for security and privacy. Federated Learning technology is adopted for scalable machine learning applications like healthcare. Furthermore, users can obtain a well-trained machine learning model without sending personal data to the cloud. Moreover, it also discussed the applications of federated learning for a distributed secure environment in a smart city.

1. Introduction

Due to digitization, a large amount of IoT data is being generated. A smart city has various use cases for AI-driven and IoT-enabled technologies, from maintaining a healthier environment to advancing advanced applications like smart healthcare. Machine learning (ML) and deep learning (DL) algorithms play a prominent role in the extraction of hidden patterns, which can be used by organizations to take business decisions based on the predictions made by these algorithms. ML/DL algorithms are being used across the domains. The data used for training the ML/DL algorithm was stored in a centralized location. The data in the central location was used to train the ML/DL algorithms. The centralized storage of data may create several problems such as security issues, single point of failure, increased latency, and others [1,2].

To overcome the drawbacks of centralized storage for machine learning, the concept of distributed learning came into the IoT environment with a smart city. In distributed learning, the training data is stored in multiple locations. The ML/DL algorithms are then fed with the data spread across several locations for training. Through distributing ML, the training time of the ML/DL algorithm decreased drastically. However, most of the problems associated with traditional ML/DL approaches persist.

Federated Learning (FL) is an extension of distributed ML that enables the ML and DL algorithms to be trained on the data located on several kinds of devices such as mobile phones, laptops, and other smart devices [3,4]. FL ensures that the computation/learning is moved to the location where the data is generated. This feature of FL improves the latency, provides more privacy and security to the data. FL also provides an additional benefit of improved personalized recommendations to users in a rapid time. Some of the popular applications of FL include personalized healthcare recommendations, movies, restaurant recommendations, etc. Apart from individual recommendations, FL has other advantages too. Even for applications that require data to be collected from multiple devices, the FL ensures that there is no need to transfer the data from the devices to the cloud. Instead, the model can be executed on the data from different devices and the parameters of the ML/DL algorithm can be updated.

Security and privacy in healthcare data are prominent challenges. Privacy is often defined as protecting sensitive information. The attack on healthcare data is due to the increased amount of medical data becoming an integral part of patient care. Authentication and encryption techniques typically prevent outsider adversaries, but the big issue comes from the insider malicious activities which cause attacks such as eavesdropping, replay attacks, denial of service attacks, and others. Table 1 summarizes the security/privacy threats of healthcare data.

* Corresponding author.
E-mail addresses: saurabh89@dongguk.edu (S. Singh), oalfarraj@ksu.edu.sa (O. Alfarraj), atolba@ksu.edu.sa (A. Tolba), postman3@dongguk.edu (B. Yoon).

https://doi.org/10.1016/j.future.2021.11.028
0167-739X/

Table 1
Threats and risk in healthcare data.

Risk type	Threats	Risk level
Privacy and trust	Eavesdropping/message disclosure	High
Integrity	Replay, Sybil, spoofing	High
Availability	DoS attack	High
Authentication	Device compromise and message alteration	High

The existing privacy-preserving mechanisms are not sufficient for full proof security of healthcare data. On the contrary, most of the medical records hosted in the cloud server are suffering from an internal attack, which is significantly much worse than the external attack. Moreover, in the medical system privacy is an important keyword for medical data as it contains sensitive information of patients. For every patient, the privacy issue varies according to the data associated with them. The important thing is how to provide appropriate privacy without loss of information.

FL is a decentralized approach that provides robust privacy-aware personalized data to several nodes within a network. It has great potential in several applications with improved performance and privacy preservation. It exploits distributed data for training and simultaneously maintains robust local learning in the local devices. Even though it has several advantages, standard FL can result in the risk of failure of a single server which is used to aggregate the learning, long communication delays [5]. In addition, an unreliable model can be uploaded by an adversary node that may lead to tampering with the learning in the FL system. Therefore, the blockchain framework, with its inbuilt features like traceability, immutability, and transparency can be integrated with FL to overcome some of the drawbacks like privacy preservation [6]. One can use the FL for the multiparty database of distributed environments. It is not only a technology standard but also an impact on the business model [7]. Federated learning trains the ML algorithms on local devices/machines/mobile phones without needing to transfer the data from the local devices to the central location like clouds. In this way, through FL, the complexity in data management and storage can be reduced. FL requires that the feature spaces shared by all participants/nodes participating are similar. However, in reality, the features shared by different nodes may be different. Moreover, the set of common entities across the devices participating in the FL may be quite small, which leaves the majority of non-overlapping data undermined and makes FL less attractive. To overcome these challenges, FL can be enhanced by transfer learning that can provide solutions for the entire sample and feature space under a federation. Since the FL trains the datasets on a local system by moving the global ML/DL algorithm to the data, several resources which would have otherwise been used to transfer the data to and from the local device to a global centralized location for training the ML algorithms. In this way, FL massively optimizes the resources.

The main characteristic of FL is that the ML/DL algorithms are executed on the data generated on a local device. Hence it supports On-Device Training. The Ge-Spatial data includes high-resolution images. On the other hand, the geospatial images are stored in the edge and then transferred to the central locations like a data center for training the ML algorithms. An alternative approach is to perform the training at the edge device itself. Although this approach has the advantage of reducing the latency, bandwidth, the geospatial data have to be still transferred from the devices to the edge which may expose the images to security risks. These problems can be solved by using FL. Related to Federated Learning for Crowd Sourcing (FLCS) are efficient. The difficulty in this is to motivate the mobile users to extend their cooperation to participate in FL. By incentivizing the mobile users who participate the FL crowdsourcing can be efficiently enhanced by FL. To summarize, the main contributions of this paper are as follows:

- Discussed security and privacy challenges of data in the healthcare system
- Blockchain and Federated Learning-enabled architecture for privacy-preservation
- Improved scalability using blockchain and Federated Learning
- Case study and performance analysis for smart healthcare

The rest of the paper is organized as follows, in Section 2, we discuss the various existing studies and the application of Federated Learning for smart applications; in Section 3, we discuss the problem formulation followed by the blockchain and Federated Learning-enabled Secure Architecture for Privacy-Preserving in Smart Healthcare in Section 4; in Section 5, we provide a case study for smart healthcare using Blockchain and Federated Learning. Finally, we conclude our paper in Section 6.

2. Related works

FL protects users from exposing their private data, while cooperatively training the global model for a variety of real-world applications. However, there are many chances that the security of FL is compromised by malicious clients. To address these issues, blockchain [8,9] can be used as a potential solution. Blockchain can be used to store ML/DL models and model parameters to improve the security of the data, realize decentralization in FL. FL can also be effectively used in blockchain technology for several improvements such as consensus efficiency of the blockchain. Moreover, mobile devices' local learning model updates can be exchanged and verified by leveraging blockchain. Blockchain has several key features, such as immutable, authentic, distributed, transparent, and decentralized [10,11]. With the use of these features in smart healthcare, we can resolve security, data integrity, centralization, and privacy-preserving. Georgios et al. have discussed the broad applications of artificial intelligence in medicine. They have discussed the requirements of privacy preservation in the healthcare system and presented an overview of current and next-generation methods for federated, secure, and privacy-preserving AI techniques focusing on medical image applications [12]. Marulli et al. proposed a security-oriented architecture for federated learning in cloud environments to improve the federated models by proposing an architecture using a blockchain scheme to get the additional features security, privacy, and scalability [13]. Li et al. [14] have reviewed the challenges and solutions for privacy-preserving through federated learning. Cheng et al. [15] proposed a new method for privacy protection in the blockchain using a trusted execution environment that guarantees the confidentiality, integrity, and authenticity of private data. Wang et al. [16] discussed the issue with traditional blockchain where the Merkle tree is not able to provide non-member proof resulting attack from a malicious node in the network. The authors have proposed a method of replacing the Merkle hash tree with a password accumulator for secure data storage. Nguyen et al. [17] proposed a blockchain chain technique to secure and sharing of IoT data across multiple data providers. This paper presents ACOMKSVM with ECC for secure and reliable data sharing. Awan et al. [18] suggested a blockchain-based privacy-preserving FL framework that is reliable and accountable that leverage decentralization of blockchain to provide provenance of model update. Rieke et al. [19] considered a key factor for the future of digital health with FL to explore the challenges and solutions that need to be addressed some more. Some more related research work on Blockchian technology and federated learning integrating with IoT and artificial intellegence can be reviewed in references [20–26].

3. Problem formulation

We have the problem for the resource management and data ownership of the blockchain enabled FL proposed framework. To achieve blockchain enabled FL, the data sample DS_j is uploaded to its associated miner. A cluster of end devices $D = \{1, 2, 3 \ldots\ldots L_d \in D\}$ where $|D| = L_d$ is achieved for block enabled federated learning. The jth device V_j possess a set of data sample DS_j where $|DS_j| = L_j \cdot D_j$. The M_j miner is selected from the set $M = \{1, 2, 3 \ldots\ldots L_M\}$ where $|M| = L_M \cdot M = D$.

Moving on, the decentralized aggregated data model training concentrate on how to solve regression problem on the parallel data sample $DS = \cup_{j=1}^{L_d} DS_j$ and $|DS| = N_P$. The nth data-sample $ds_n \in DS$. Now, the problem is to minimize loss function of data ownership $f(\alpha)$ in respect with dimensional vector $\alpha \in R^d$ where d is dimensional column parameter. So, to simplify, the predefine loss function for data processing is selected as mean squared error through the following Eq. (1):

$$f(\alpha) = \frac{1}{L_D P} \sum_{j=1}^{L_d} \sum_{p_n} f_n(\alpha) \tag{1}$$

where $f_n(\alpha)(x_n^T \alpha - y_n)^2 \div 2$ and the $(.)^T$ notation denotes vector transpose operation. Although, the work proposed in the article considers privacy as a core issue, this data processing forms a key part when FL must be applied along with the proposed architecture. For the sake of simplicity, we consider that the architecture receives clean data for sharing, and a data-exchange protocol helps to amalgamate the services for the user device using both FL and the blockchain.

4. Proposed architecture for smart healthcare

4.1. Architecture overview

Fig. 1 depicts the abstract system architecture of a general IoT blockchain cloud platform. The system comprises three main components: the sensor network from where the private data centers are stored by the data publishers through the IoT devices presented at the user level and sent to the blockchain cloud network where data is validated and processed and delivered to a third component such as monitors, healthcare equipment like MRI machines, smartphones, measuring devices for a heartbeat and another phenomenon, and monitors. These devices are aimed at the acquisition of data from the body like temperature, heartbeat, blood pressure, electrocardiograph and many others.

The rest web service which is a lightweight web interface is used by sensors for sending data and requested data from the cloud. For the faster management of data ubiquitously with a variety of interfaces (i.e., PCs, TVs, smartphones), the cloud server provides powerful machines like high CPU, computation power, memory, GPU and sufficient network bandwidth as per demand. The Blockchain cloud layer is present at the control layer where the public data is responsible for secure communication and computation. For data aggregation and collaboration, the blockchain system records the data usage behavior and ensures authenticity as well. When the data user uses the target data, it is handled by our system which is then processed and give the result that returns to the data user. It enables data ownership separation and permission that makes the data rotation process safe and controllable. The quality of training data strongly impacts the effectiveness of the model by assigning the labels to training data items. The data labeling process is to identify raw data and add some meaningful labels to provide context so that the FL model can learn from that and compile it. The way of assigning the labels are: (a) add data items to a dataset with labeled assigned, (b) using console assign labels to data items, (c) human labelers add labels to data items. Unified data learning of FL for multilateral computing is completed through the public data center. FL takes in data calculation instead of considering data itself which results from the data security of users while data remain present in the private data center. computational logic provides techniques for analyzing the interfering properties of language with high-level analysis and implementations. And the data analyzer analyzes current and historical healthcare data to predict trends lines data and improve outreach.

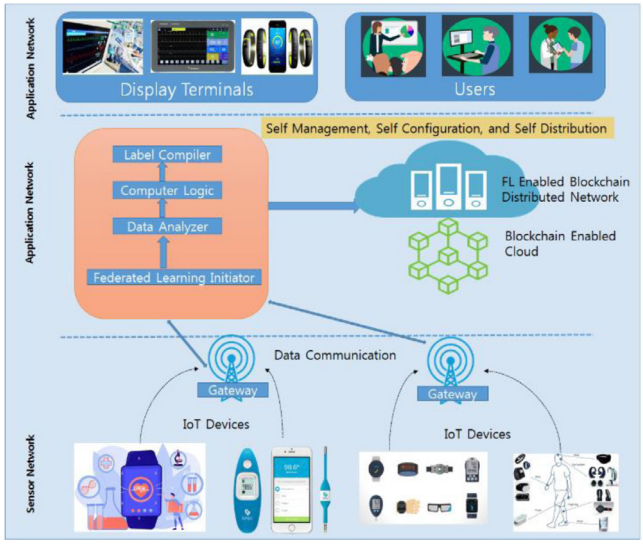

Fig. 1. Overview of the proposed framework and hierarchy.

4.2. Reward function

To achieve certain model accuracy and reduce energy consumption and training latency with fair payment system. The immediate reward in the system can be calculated by the following Eq. (2).

$$R(S_t A_t) = \beta_Q \frac{Q}{Q_{max}} - \beta_S \frac{S}{S_{max}} - \beta_T \frac{T}{T_{max}} - \beta_U \frac{U}{U_{max}} \tag{2}$$

where β_Q, β_S, β_T, and β_U are scale factors, Q is the data unit, and β is the action $\beta_t \in A$ taken which receive immediate rewards. S is energy unit consumed by the device, T is total latency and U is final payment added by cost of devices paid for training and miner wining for block generation in chain.

Blockchain data registration is aimed to configure registered data uploaded by the private data center to check data ownership in the healthcare system. And the blockchain system is aimed to configure data management which has an owner relationship with the data editor. It also trusts on configured Blockchain to record the collaborative and transactional behavior. It can also share the monitored data with authorized healthcare communities to search for personalized trends and patterns. The communication in the network is secured by an authentication scheme through cryptographic techniques of healthcare data. The proposed architecture has also facilitated building off a context-aware sensing capability, cost-effective, autonomous, and health management at any time and place.

The distributed blockchain cloud network protects patient information, offers precise solutions of trust and security issues, increases productivity with high availability by resource providing. The network also provides independent large storage for

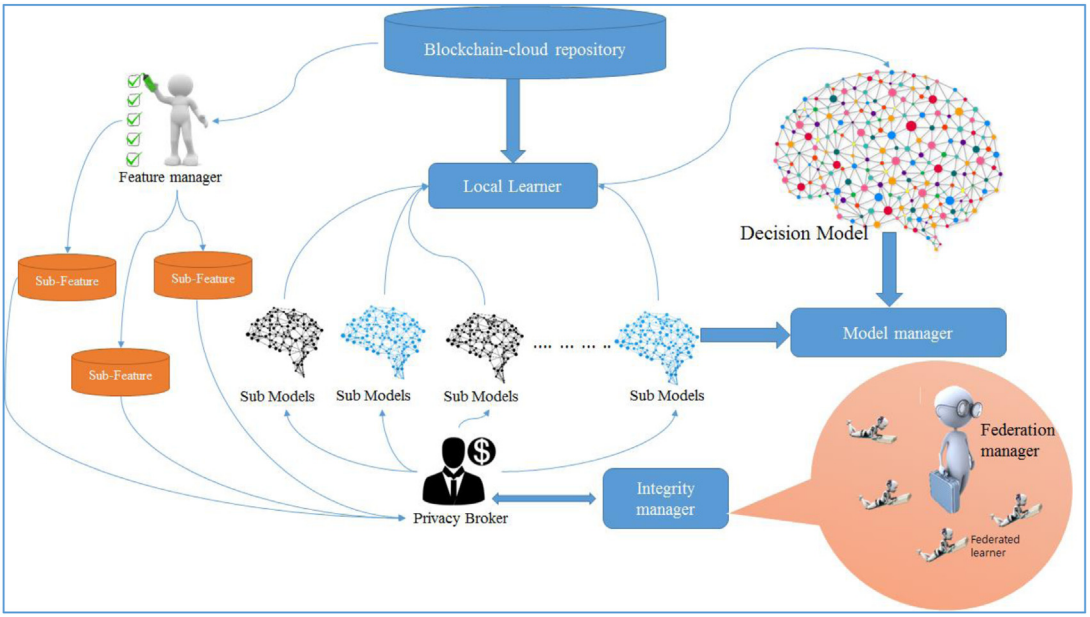

Fig. 2. Working methodological flow of the proposed architecture.

healthcare data. To protect the crucial data in the healthcare system it uses cryptographic techniques such as encryption, hashing, and signature.

4.3. Federated working architecture

Fig. 2 discusses the detailed description of privacy protection mechanisms that define hierarchical predictors, where different sub-models are trained on sub-feature—gesture, voice, handwritten and others. The aggregated data are vertically parted in different data sets accomplished by the sub-feature manager. In that, the partitioning mechanism is categorized as per federated learning feature, it is becoming more useful tooling AI-based applications. In the federation phase, every sub-feature dataset is used to help for training the different sub-models through federations. The client or user provides data to the federation manager and received a sub-model to train it. This coalition process is intermediated by the privacy broker that is guaranteed to resolve the issue of privacy. Another component is the integrity manager that maintains the result integrity by avoiding errors inside sub-models. FL reduces the risk of sensitive healthcare data to address the privacy issue. Moreover, secure multiparty computation has an application to FL scenarios where each node has cryptographic combination and oblivious transfer to jointly compute a function of their private data. In this way, the FL algorithm reduces the risk of exposing healthcare patient data and, while dealing with medical records blockchain addresses transparency and user consent.

4.4. Data communication

The data center is connected to a blockchain and federated learning-based network where it manages the data and generates data units that are easy to share in the network and confirm ownership. The natural language processing (NLP) is subjected to configure inbuilt in the network to process the data which is uploaded by data publisher to data generator unit. Data registration generators ensure each data in the system has a unique digital identity that tracks, collaborates, and circulates properties of data assets.

4.5. Privacy preserving

Optimizing the blockchain-based federated learning for the proposed system architecture requires some important considerations. A consortium can contain limited participants than a public network, but that each participant can have many on-device samples. Another thing is that these participants have more resources connected to the network. These connected devices satisfy almost all criteria. Moreover, privacy in FL is reflected by reputed academic research scientist's work in this area to protect the training data. To provide security and privacy in the blockchain, a secure ledger consist of transaction records in a chain of blocks. Each block is guarded by cryptographic techniques enforcing integrity in the records. Moreover, each block maintains the hash value of the previous block that serves as a cryptographic linkage to its preceding block. The proof-of-work blockchain consensus protocol makes the blockchain trustworthy enforced by the network which controls the generation and admission of the new block, verification and validation by the proof-of-work consensus protocol of blockchain and the copy of record ledger available to each node that maintains transparency in blockchain

Concerning the very large size of data, an inversion or reconstruction attack can cause catastrophic data leakage, which turns out to reconstruct the image with more visualization than the original training data with impressive accuracy and details. Therefore, federated learning provides an infrastructure approach to privacy and security with the help of differential privacy and homomorphic encryption. In addition, the role of homomorphic encryption is public-key encryption where a node encrypts its data and performs calculations with data encrypted by others with the same public key. With its success in cloud computing, it naturally entered this area and is certainly used in many federated learning studies. Differential privacy is an alternative theoretical model for protecting privacy that anyone seeing the result of differentially private analysis of individual data. As widely studied in FL, with the help of deep learning, support vector machine and principle component analysis, differential privacy ensures that addition or removal does not significantly affect the outcome of the analysis.

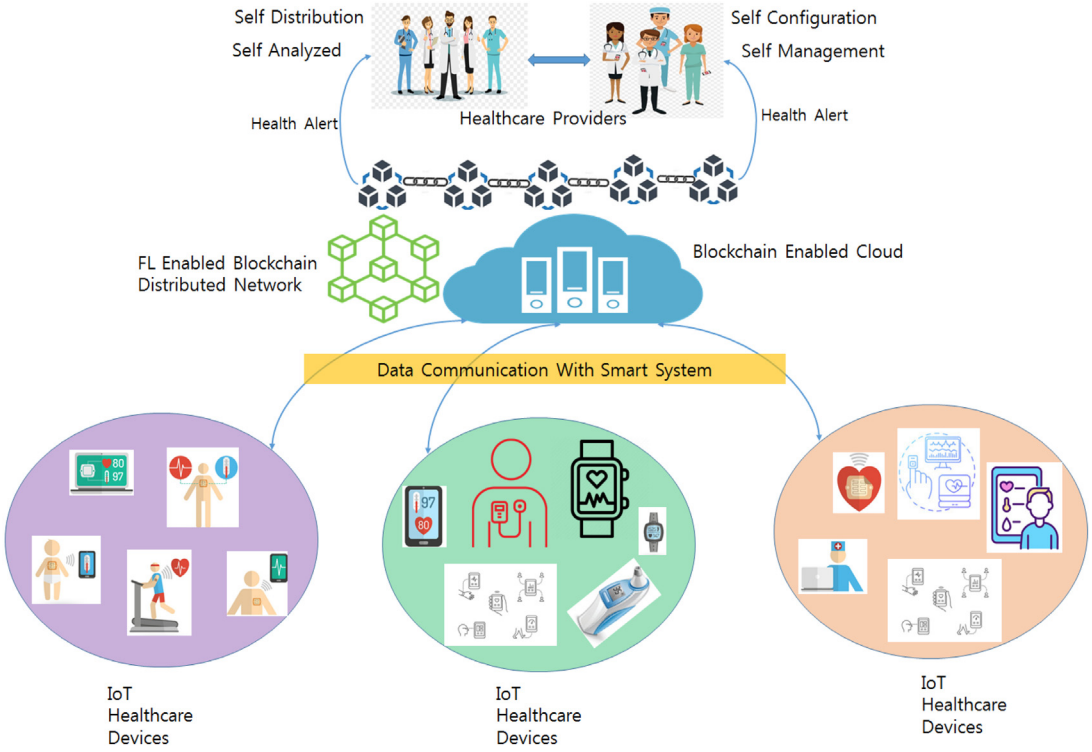

Fig. 3. Case study on health care using the proposed framework.

5. Case study for smart healthcare

In this section, we discuss a case study that is based on the proposed architecture. IoT devices, including weight meters, blood pressure, glucose meter, insulin pump, and others are connected to patients [27]. It generates raw data reports and transfers them to IoT devices. These devices communicate the data to smart systems such as a smart monitor, laptops, and mobiles. These systems have functions, including translation, compression, and formation. Then this data transfers to the Blockchain and Federated learning-based distributed network (Blockchain-enabled cloud). Blockchain provides security and privacy in a distributed way and stores in a tamper-proof ledger with the help of some protocols of blockchain technology. It verifies and validates by the miner nodes in the network. Federated learning analyzes the data, if it has a problem, then sends the alert to the smart system, health alert transfer to the health care provider, and does not add a block in the blockchain network. If data is alright, it transfers to the healthcare provider, which manages and prescribes the medicine to the patient with some functions, including self-configuration, self-distribution, and self-management. This case study provides the privacy-preserving by blockchain and federated learning to smart healthcare. An exemplary overview of the case study is shown in Fig. 3, and the protocol procedures are given in Fig. 4.

5.1. Understanding the security of the system

The security of the healthcare system using the proposed architecture is defined in terms of its ability to handle privacy issues without affecting the flow of the system. Here, theoretical evaluations are used to understand the impact on the security of the system as defined below:

Remark 1. The privacy of the system is attained through data rotation over federated learning.

Claim. *In the proposed model, the federated learning is performed using public data centers, which also helps to differentiate between the permissions for the owner which controls the safety of the data-rotation. Thus, maintaining the privacy of the system.*

Remark 2. The initial phase of the proposed model ensures the security of the healthcare data of the patients

Claim. *In the initial phase of the system, the blockchain and the federated learning-based distributed network helps to maintain the flow of content through the sub models, which are operated through the privacy broker system which exchanges the control messages with the integrity manager to ensure that the policy updates are no shared with the global server and only the settings are available to manage the flow over the deployed architecture. The healthcare provider receives alerts on the analyzed content from the smart system whereas the data of the patient is not in its peripheral system. Thereby, securing the entire model.*

Remark 3. Security properties can be enhanced through protocols operating between the patient-side devices and the healthcare provider and the security is impacted by the number of passes between the modules.

Claim. *In the proposed healthcare case study over the proposed architecture, the security can be enhanced by incorporating protocols on the IoT devices and the smart system which can exchange the initial control messages to ensure features of message authentication, perfect-forward secrecy and prevention against replay attack using which an eavesdropper may try to violate the privacy of the patient. However, the cost of deploying these protocols will impact the performance of the system, which is specifically affected by the number of passes the protocol must follow. Alongside this, there needs to have end to end security which will require additional protocols between the smart system and the healthcare providers. However, considering the federated learning and blockchain framework are secure under a*

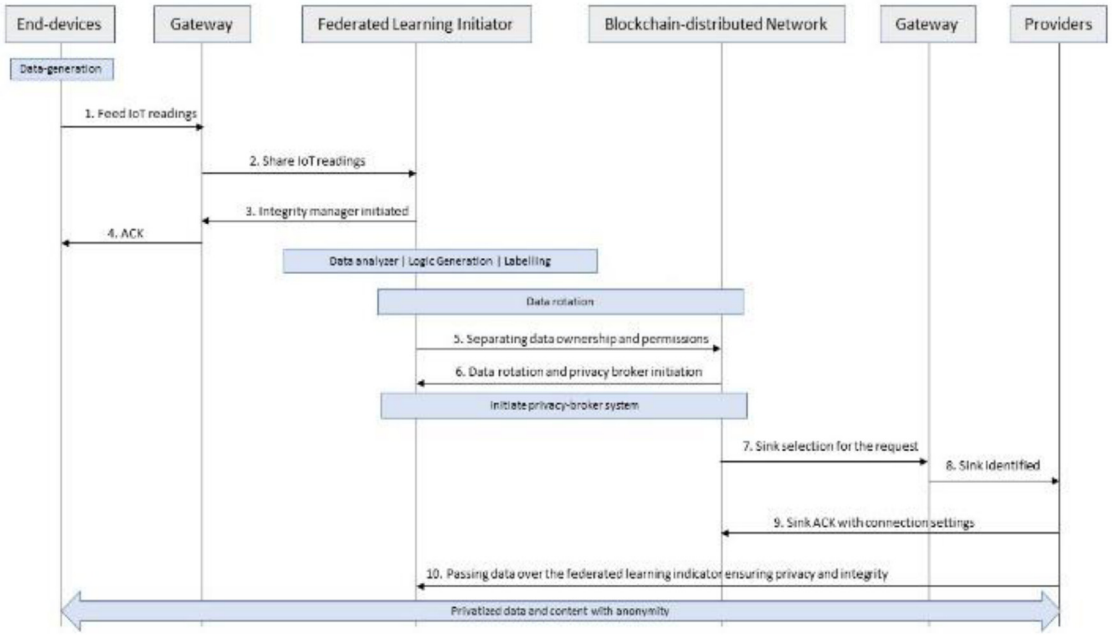

Fig. 4. Protocol procedures for the case study over the proposed framework.

given consensus model, such end-to-end security can be embedded within the blockchain through smart contracts. Alternatively, the model can be enhanced with the Proof of Integrity, which will ensure that the system has one set of protocols operating only between the IoT-end devices and the integrity manager, whereas the blockchain protocol ensures manages the provider side security.

5.2. Understanding the performance of the system

To understand the performance of the system, a numerical modeling is considered which combines the federated learning model, blockchain model and the communication protocol. The impact of the system is followed in three main parts.

The first one covers the evaluation of the number of passes the workflow needs to follow if additional security protocols are used. The second evaluates the scalability of the system using the probabilistic situation where the local updates are modeled with Poisson Distribution, and the third one presents the reliability based on the evaluation of the number of passes and the scalability.

For the number of passes, the modeling in [28] is used to identify the overheads of the messages which are then incorporated to show the impact of the passes on the performance of the system. In the considered case study, the overheads with the number of passes are calculated as:

$$O_s = \sum_{i=1}^{N} (\tau(R_l + Pr.R_H))_i \tag{3}$$

Here, N is the number of entities passing information between them. τ is the time to pass the information between the entities with one-hop distance, Pr is the probability of a service being allocated over the blockchain and federated learning based distributed network, R_l is the incoming requests from end devices near the patients, R_H is the requests handled by the smart system at a given instance.

Fig. 5 discusses the results associated with the workflow of the case study on healthcare using numerical case studies. The overheads are impacted by the delays associated with the network. Specifically, blockchain and training delays can heavily impact

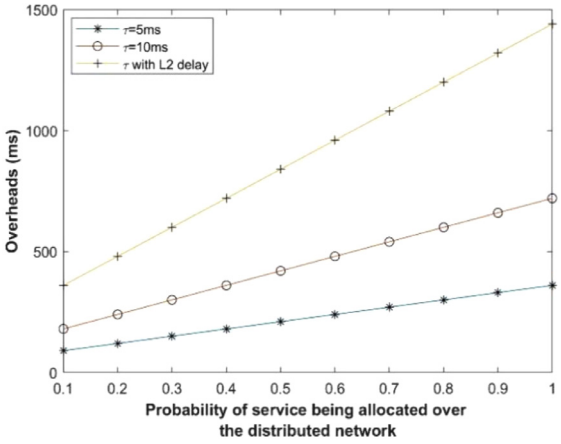

Fig. 5. Overheads vs. probability of a service being allocated to the blockchain-federated learning based distributed network at a fixed probability of support from the infrastructure.

any system using them. However, there is a choice to make and settle the trade-off between privacy and the performance. The results in Fig. 5 consider general delays between 5 and 10 ms as well as L2 delay of 20 ms caused due to network's processing. The number of entities required to perform the actions is 6 between the patient and the healthcare provider where it is assumed that each intermediate entity in the infrastructure can accommodate 10 service requests simultaneously and new requests arrive at a rate of 2 per second. The overheads are impacted by the probability of services being allocated to a more distributed network. More requests mean more computational load on the network. With the given model, increasing probability, which means more handling of the services still make lower overheads, making the approach efficient along with privacy-preserving. It is to be noticed that different connectivity for the entities in the model will impact the overheads differently.

Some devices may pose high service probability compared to others, which impacts the performance in terms of overhead as shown in Fig. 6. With lower probabilities, the network utilization

Table 2
Comparison with the state-of-the-art solutions.

Article	Ideology	Privacy-preservation	Blockchain	Overheads	Blockchain + FL for scalability
Puri et al. [29]	Decentralized healthcare	No	Yes	Medium (transactions overhead)	No
Tanwar et al. [30]	Fog-based healthcare classifier	No	No	Low	No
Abou-Nassar et al. [25]	Blockchain based trust for healthcare	Yes	Yes	–	No
Guo et al. [26]	Data clustering for IoT-based healthcare	Yes	No	–	No
Proposed Model	Scalable, decentralized blockchain and FL based healthcare	Yes	Yes	Low	Yes

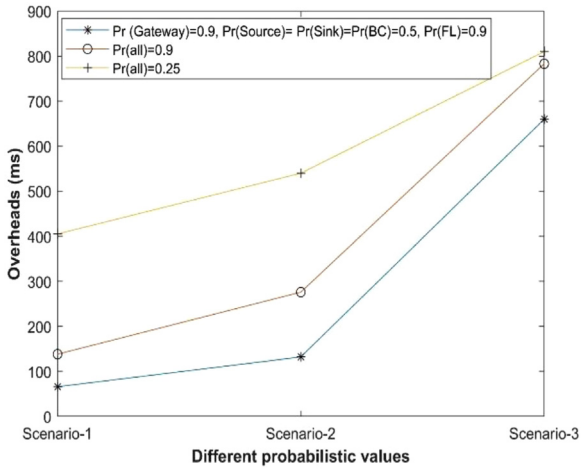

Fig. 6. Overheads vs. probability of a service being allocated to the blockchain-federated learning based distributed network with a varying probability of support from the infrastructure.

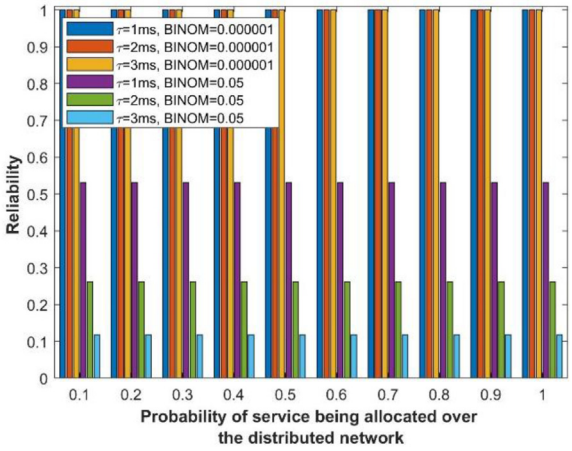

Fig. 7. Reliability vs. probability of a service being allocated to the blockchain-federated learning based distributed network with varying β and τ.

is less, which results in lower overhead that means privacy is maintained but connections may not be available at all instances. Such situations need to be considered when using the proposed framework under live settings.

The reliability of the system is evaluated with a varying time and the number of entities with respect to the probability distribution, which considers the failure and inclusion of the new entities. In the given model, the reliability is calculated over binomial distribution of connections, using the model in [31,32]. According to which,

$$R_e = \prod_{i=1}^{S}(1 - \sum_{j=1}^{M} \tau_j(\prod_{k=1}^{B}(\binom{R_{T,i,k}}{R_H}\beta_{i,k}{}^{R_H}(1 - \beta_{i,k}{}^{R_{T,i,k}-R_H}))) \quad (4)$$

Here, S is the total subnetworks equivalent to N, M is the number of smart systems, B is the blockchain and federated distributed network offering alternative routes, R_T is the total number of requests over the model, β is the probability defining the failure of new connection requests.

In variable scenarios, where the delays are dominant, it is taken as a constant with a minimum value as $\frac{1}{0_s}$.

In the model, the reliability is accounted based on the overheads and procedures the patient's end-device needs to follow to get observation outcomes from the healthcare provider. Based on the observations and using (4), results are discussed in Fig. 7. These numerical results suggest that with an increasing probability of allocating services, the reliability is much impacted by the processing time at each entity that adds to the delay. In the evaluations, the delays at each device were set between 1 and 3 ms where two different cases of β were considered. The first case is observed based on the overheads whereas in the second case, the values of the binomial distributions are prefixed to understand the impact on the reliability based on the variations in τ.

Furthermore, a qualitative comparison is given against the approaches which have a similar theme and rely on IoT devices for healthcare applications, as shown in Table 2. Existing techniques have used blockchain to handle healthcare privacy but have not exploited it with FL to improve scalability and privacy. It makes our solution different. Moreover, our model is checked against overheads, whereas existing studies have not quantitatively measured the overheads in similar settings.

At present, the proposed model is limited to a theoretical model, but has the capacity to be applied to a live blockchain and federated learning-based distributed network, which will impact the performance because of the additional delays. However, the model is applicable considering the advantages of the privacy preservation of healthcare data.

6. Conclusion

Federated learning revolutionizes the way of machine learning by providing privacy-preserving, security, and scalability. The proposed federated based blockchain cloud architecture has provided secure data collaboration for the IoT environment. FL training mechanism where the model with client–server architecture keeping the sensitive data to the local bodies, provides healthcare data source with privacy preservation. And it is lightweight, scalable and supports interoperability. Cloud computing service providers extend the benefits of healthcare users by using efficient broker management. And, a peer-to-peer network provided by Blockchain technology that supports communication between non-trustable nodes with a scalable processing network. Currently, this work relies on the performance of the protocol used for amalgamating Blockchain and FL to support the privacy of the end device. However, this limitation of only relying on the protocol will be resolved by building a trust model based on blockchain and having a relatively new consensus mechanism for supporting FL nodes. In the future, we aim to optimize the latency

and storage requirements. We will also devise the level of trust and reward mechanism for the user device.

CRediT authorship contribution statement

Saurabh Singh: Conceptualization, Methodology, Software, Data curation, Writing - original draft. **Shailendra Rathore:** Visualization, Formal Analysis, Investigation, Supervision, Software, Validation, Writing - review & editing. **Osama Alfarraj:** Conceptualization, Methodology, Software, Data curation, Writing - original draft. **Amr Tolba:** Conceptualization, Methodology, Software, Data curation, Writing - original draft. **Byungun Yoon:** Visualization, Formal Analysis, Investigation, Supervision, Software, Validation, Writing - review & editing.

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.

Acknowledgments

This work was supported in part by the National Research Foundation of Korea under Grant 2019R1A2C1085388. This work was funded by the Researchers Supporting Project No. (RSP-2021/102) King Saud University, Riyadh, Saudi Arabia. We would like to thank you to Dr. Vishal Sharma from Queen's University Belfast, UK for helping in paper contribution also thank to Mr. Sushil Kumar Singh from Seoultech University, South Korea.

References

[1] Ethem Alpaydin, Introduction to Machine Learning, MIT Press, 2020.
[2] V. Sharma, I. You, K. Andersson, F. Palmieri, M.H. Rehmani, J. Lim, Security, privacy and trust for smart mobile-internet of things (M-IoT): A survey, IEEE Access 8 (2020) 167123-167163.
[3] W.Y.B. Lim, N.C. Luong, D.T. Hoang, Y. Jiao, Y.C. Liang, Q. Yang, D. Niyato, C. Miao, Federated learning in mobile edge networks: A comprehensive survey, IEEE Commun. Surv. Tutor. (2020).
[4] Qiang Yang, Yang Liu, Tianjian Chen, Yongxin Tong, Federated machine learning: Concept and applications, ACM Trans. Intell. Syst. Technol. (TIST) 10 (2) (2019) 1–19.
[5] G.A. Kaissis, M.R. Makowski, D. Rückert, R.F. Braren, Secure, privacy-preserving and federated machine learning in medical imaging, Nat. Mach. Intell. (2020) 1–7.
[6] Fiammetta Marulli, Emanuele Bellini, Stefano Marrone, A security-oriented architecture for federated learning in cloud environments, in: Workshops of the International Conference on Advanced Information Networking and Applications, Springer, Cham, 2020, pp. 730–741.
[7] Z. Li, V. Sharma, S.P. Mohanty, Preserving data privacy via federated learning: Challenges and solutions, IEEE Consumer Electron. Mag. 9 (3) (2020) 8–16.
[8] S. Singh, P.K. Sharma, B. Yoon, M. Shojafar, G.H. Cho, I.H. Ra, Convergence of Blockchain and Artificial Intelligence in IoT Network for the Sustainable Smart City, Sustainable Cities and Society, 2020, p. 102364.
[9] Bordel Borja, Alcarria Ramon, Martin Diego, Sanchez Picot Alvaro, Trust provision in the internet of things using transversal blockchain networks, Intell. Autom. Soft Comput. 25 (1) (2019) 155–170.
[10] V. Sharma, I. You, F. Palmieri, D.N.K. Jayakody, J. Li, Secure and energy-efficient handover in fog networks using blockchain-based DMM, IEEE Commun. Mag. 56 (5) (2018) 22–31.
[11] Xin Jiang, Mingzhe Liu, Chen Yang, Yanhua Liu, Ruili Wang, A blockchain-based authentication protocol for WLAN mesh security access, Comput. Mater. Contin. 58 (1) (2019) 45–59.
[12] A. Bhowmick, J. Duchi, J. Freudiger, G. Kapoor, R. Rogers, Protection against reconstruction and its applications in private federated learning, 2018, arXiv preprint arXiv:1812.00984.
[13] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, Y. Zhou, A hybrid approach to privacy-preserving federated learning. In Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security (1–11), 2019.
[14] Qiang Yang, Yang Liu, Tianjian Chen, Yongxin Tong, Federated machine learning: Concept and applications, ACM Trans. Intell. Syst. Technol. (TIST) 10 (2) (2019) 1–19.
[15] Jin Wang, Wencheng Chen, Lei Wang, R. Simon Sherratt, Osama Alfarraj, Amr Tolba, Data secure storage mechanism of sensor networks based on blockchain, CMC-Comput. Mater. Contin. 65 (3) (2020) 2365–2384.
[16] Sharma Pradip Kumar, Jong Hyuk Park, Kyungeun Cho, Blockchain and federated learning-based distributed computing defence framework for sustainable society, Sustainable Cities Soc. 59 (2020) 102220.
[17] Le Nguyen Bao, E. Laxmi Lydia, Mohamed Elhoseny, Irina Pustokhina, Denis A. Pustokhin, Mahmoud Mohamed Selim, Gia Nhu Nguyen, K. Shankar, Privacy preserving blockchain technique to achieve secure and reliable sharing of IoT data, Comput. Mater. Contin. 65 (1) (2020) 87–107.
[18] Sana Awan, Fengjun Li, Bo Luo, Mei Liu, Poster: A reliable and accountable privacy-preserving federated learning framework using the blockchain. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, 2561–2563. 2019.
[19] Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R. Roth, Shadi Albarqouni, Spyridon Bakas, et al., The future of digital health with federated learning, NPJ Digital Med. 3 (1) (2020) 1–7.
[20] Saurabh Singh, Pradip Kumar Sharma, Byungun Yoon, Mohammad Shojafar, Gi Hwan Cho, In-Ho Ra, Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city, Sustainable Cities Soc. 63 (2020) 102364.
[21] Jieren Cheng, Jun Li, Naixue Xiong, Meizhu Chen, Hao Guo, Xinzhi Yao, Lightweight mobile clients privacy protection using trusted execution environments for blockchain, CMC-Comput. Mater. Contin. 65 (3) (2020) 2247–2262.
[22] Yongjun Ren, Yan Leng, Jian Qi, Pradip Kumar Sharma, Jin Wang, Zafer Almakhadmeh, Amr Tolba, Multiple cloud storage mechanism based on blockchain in smart homes, Future Gener. Comput. Syst. 115 (2021) 304–313.
[23] Jieren Cheng, Jun Li, Naixue Xiong, Meizhu Chen, Hao Guo, Xinzhi Yao, Lightweight mobile clients privacy protection using trusted execution environments for blockchain, CMC-Comput. Mater. Contin. 65 (3) (2020) 2247–2262.
[24] Bo Yin, Hao Yin, Yulei Wu, Zexun Jiang, FDC: A secure federated deep learning mechanism for data collaborations in the internet of things, IEEE Internet Things J. 7 (7) (2020) 6348–6359.
[25] E.M. Abou-Nassar, A.M. Iliyasu, P.M. El-Kafrawy, O.Y. Song, A.K. Bashir, A.A. Abd El-Latif, DITrust chain: towards blockchain-based trust models for sustainable healthcare IoT systems, IEEE Access 8 (2020) 111223-111238.
[26] X. Guo, H. Lin, Y. Wu, M. Peng, A new data clustering strategy for enhancing mutual privacy in healthcare IoT systems, Future Gener. Comput. Syst. 113 (2020) 407–417.
[27] A.D. Dwivedi, G. Srivastava, S. Dhar, R. Singh, A decentralized privacy-preserving healthcare blockchain for IoT, Sensors 19 (2) (2019) 326.
[28] S.M. Razavi, D. Yuan, F. Gunnarsson, J. Moe, Exploiting tracking area list for improving signaling overhead in LTE, in: 2010 IEEE 71st Vehicular Technology Conference, IEEE, 2010, pp. 1–5.
[29] V. Puri, A. Kataria, V. Sharma, Artificial intelligence-powered decentralized framework for internet of things in healthcare 4.0, Trans. Emerg. Telecommun. Technol. (e4245) (2021).
[30] S. Tanwar, J. Vora, S. Kaneriya, S. Tyagi, N. Kumar, V. Sharma, I. You, Human arthritis analysis in fog computing environment using Bayesian network classifier and thread protocol, IEEE Consum. Electron. Mag. 9 (1) (2019) 88–94.
[31] V. Sharma, I. You, D.N. Jayakody, D.G. Reina, K.K. Choo, Neural-blockchain-based ultrareliable caching for edge-enabled UAV networks, IEEE Trans. Ind. Inform. 15 (10) (2019) 5723–5736.
[32] D.W. Coit, A.E. Smith, Reliability optimization of series-parallel systems using a genetic algorithm, IEEE Trans. Reliab. 45 (2) (1996) 254–260.

Saurabh Singh received the Ph.D. degree from Jeonbuk National University, Jeonju, South Korea, carrying out his research in the field of ubiquitous security. He was a Postdoctoral Researcher with Kunsan National University, South Korea. He currently joined as an Assistant Professor with Dongguk University, Seoul, South Korea. He has published many SCI/SCIE journals and conference papers. His research interests include blockchain technology, cloud computing and security, the IoT, deep learning, and cryptography. He received the Best Paper Award from KIPS and CUTE Conference, in 2016.

Dr. Shailendra Rathore is working as a Lecturer of cyber security in the School of Engineering, Computing and Mathematics at University of Plymouth, UK. He received Ph.D. degrees in the Graduate School of Computer Science and Engineering at Seoul National University of Science and Technology, Seoul, South Korea. His broad research interest includes Cyber Security, Artificial Intelligence, Blockchain, and Machine Learning. Before joining Ph.D., Dr. Rathore has worked as an Executive – Technology at Crompton Greaves Global R & D, Mumbai, India from June 2013 to July 2014. He had been a research and teaching assistant at the Department of Computer Science and Engineering, National Institute of Technology, Kurukshetra, India. Dr. Rathore received his M.E. in Information Security from Thapar University, Patiala, India (July 2014), and B.Tech. in Computer Engineering from Rajasthan Technical University, Kota, Rajasthan, India (July 2012). Dr. Rathore has published his research outcomes in various top tier international journals, including IEEE communication magazine, IEEE network, IEEE consumer electronics, FGCS, Information sciences, IEEE access He has been serving as chair, program committee, or organizing committee chair for many international conferences and workshops, including IEEE, ACM.

Osama Alfarraj received the master's and Ph.D. degrees in information and communication technology from Griffith University, in 2008 and 2013, respectively. He is currently an Associate Professor of computer sciences with King Saudi University, Riyadh, Saudi Arabia. His current research interests include eSystems (eGov, eHealth, and ecommerce), cloud computing, and big data. He has served as a Consultant and a member of the Saudi National Team for Measuring E-Government, Saudi Arabia, for two years. (oalfarraj@ksu.edu.sa).

Amr Tolba (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from Mathematics and Computer Science Department, faculty of science, Menoufia University, Egypt, in 2002 and 2006, respectively. He is currently an Associate Professor at the Faculty of Science, Menoufia University, Egypt. He is currently on leave from Menoufia University to the Computer Science Department, Community College, King Saud University (KSU), Saudi Arabia. Dr Tolba serves as a technical program committee (TPC) member in several conferences. He has authored/coauthored over 100 scientific papers in top ranked (ISI) international journals and conference proceedings. His main research interests include artificial intelligence (AI), the Internet of Things (IoT), data science, and cloud computing. (atolba@ksu.edu.sa).

Byungun Yoon (Senior Member, IEEE) is currently a Professor with the Department of Industrial and Systems Engineering, Dongguk University. His theme of study has involved blockchain technology, patent analysis, new technology development methodology, and visualization algorithms. His current research interests include enhancing technology road mapping, research and development quality, and product designing with data mining techniques