Blockchain for Healthcare: Securing Patient Data and Enabling Trusted Artificial Intelligence

H.S. Jennath[1]*, V.S. Anoop[2], S. Asharaf[3]

[1] Data Engineering Lab, Indian Institute of Information Technology and Management – Kerala (IIITM-K), Thiruvananthapuram, 695581 (India)
[2] Kerala Blockchain Academy (KBA), Indian Institute of Information Technology and Management – Kerala (IIITM-K), Thiruvananthapuram, 695581 (India)
[3] Indian Institute of Information Technology and Management – Kerala (IIITM-K), Thiruvananthapuram, 695581 (India)

Received 10 December 2019 | Accepted 11 January 2020 | Published 2 July 2020

unir
LA UNIVERSIDAD
EN INTERNET

Abstract

Advances in information technology are digitizing the healthcare domain with the aim of improved medical services, diagnostics, continuous monitoring using wearables, etc., at reduced costs. This digitization improves the ease of computation, storage and access of medical records which enables better treatment experiences for patients. However, it comes with a risk of cyber attacks and security and privacy concerns on this digital data. In this work, we propose a Blockchain based solution for healthcare records to address the security and privacy concerns which are currently not present in existing e-Health systems. This work also explores the potential of building trusted Artificial Intelligence models over Blockchain in e-Health, where a transparent platform for consent-based data sharing is designed. Provenance of the consent of individuals and traceability of data sources used for building and training the AI model is captured in an immutable distributed data store. The audit trail of the data access captured using Blockchain provides the data owner to understand the exposure of the data. It also helps the user to understand the revenue models that could be built on top of this framework for commercial data sharing to build trusted AI models.

Keywords

Blockchain, Healthcare, Data Privacy And Security, Trusted Artificial Intelligence, Interoperability, Patient-Centric Data Management, Provenance, Traceability.

DOI: 10.9781/ijimai.2020.07.002

I. Introduction

HEALTHCARE data is very sensitive due to the inclusion of personal information, hence this industry demands requirements like security and privacy of storage and access of data, confining to legal obligations in protecting their health information [1]. Advances in information technology(IT) enabled the digitization of the health records which further made the data sharing across the stakeholders much easier. Traditionally healthcare datarests within the vicinity of the healthcare provider. These organizations usually manage their in house data of their patients using healthcare customer relationship management systems or electronic health records(EHR). EHR permits digital data capture as well as storage of medical records and access to authorized users [15]. EHRs can assist doctors or healthcare professionals with decision making and can provide insight into the way a patient is treated. Recent accelerated digitization adoption in healthcare aims to provide better patient care, with more accurate analysis and diagnosis and secured access to medical data on demand at a reduced cost. This is achieved by avoiding multiple diagnoses or redundant drug administration, etc. With the advent of Artificial Intelligence (AI), the momentum of innovation in digitization in healthcare is gaining momentum and the industry is now all set to accelerate further with Blockchain technology [18][19].

Advances in the area of wireless communication, distributed storage and edge computing, paved the path to accelerated development and deployments of wearable devices and cloud storage for ubiquitous monitoring and recording of health-related information. The sensitive data stored and processed in e-health systems are mandated to have adopted enough security and privacy measures to prevent disclosure or leakage of medical data as it causes a significant impact on a patient's social life. Leaked medical info would cause a lot of problems in patients personal life such as a loan request rejection, surge in insurance premium, rejection of job application, etc. Hence any deliberate or accidental data breach of healthcare records violating the privacy or confidentiality concerns of patients would lead to severely penalizing the healthcare provider or the IT (Information Technology) infrastructure provider [16], [17].

Mobile Health (mHealth) is another prominent field in the healthcare industry that employs IoT devices or wearable sensors, low-power body-area wireless networks, and smart devices [16], [17], [18]. One of the major challenge associated with healthcare domain is the data access through smart devices. There is an increased chance of malicious attacks or compromised security through these channels. Moreover, when patients are consulting different doctors or various healthcare providers, the sharing of this sensitive information and its privacy and confidentiality is a concern. Hence, while managing the Healthcare data, the multitude of challenges needs to be addressed. The major challenges include but not limited to the authentication of access permission, interoperability of health records, data sharing policies

* Corresponding author.

E-mail address: jennath.res16@iiitmk.ac.in

or rule enforcement on shared content, the secured sharing of health records, and security and privacy concerns about wearables and smart devices for mobile health data capture and processing.

Interoperability in healthcare has been an exciting area of interest for a long time regarding data exchanges between various stakeholders, like insurance providers, researchers, different health care providers, etc. However, this has caused another implication in securing the privacy and identity concerns of the patient. A recent trend in healthcare data management is in a transition phase from organization centric data management towards patient-driven, patient-mediated data management, as patients are true owners of the data. Recent privacy regulations like the General Data Protection Regulation (GDPR) and HIPAA (Health Insurance Portability and Accountability Act of 1996) also demanded a patient mediated or patient-driven data interoperability. In short, the concern mainly revolves around the implementation of a patient-driven data sharing mechanism conserving his privacy, security, and audit-ability of the shared data [9]. However, there exists some ambiguity regarding whether patients should be able to access their entire health data. HIPAA recommends that every individual has the right to access their health data on demand (excluding exceptions, like notes on psychotherapy) [20]. Blockchain is a distributed ledger technology that provides a trusted infrastructure platform for setting up a multi-party business network for mutually distrusting agents. The immutability of records in the shared ledger is achieved through the hashed linking of blocks and the efficient consensus algorithm which makes sure the persisted data is foolproof. All the participants in the network have the same copy of ledger with the same data. The Blockchain mandates the business contracts across the participating parties in the network using the powerful self-executing smart contact that enforces the contractual agreement [21][31].

In this work, we explore the potential of blockchain in leveraging a patient-centric data interoperability architecture for medical records sharing, where data owner decides the granularity of information to be disclosed with intended parties. This work explores the hierarchical architecture using blockchain which captures the traceability of the data access policies set by the data owner, an audit trail of the details of data accessed by third party requesters, incentive agreements with the third-parties for the data sharing, etc. This work also explores the potential of building trusted Artificial intelligent models using e-health data. AI or machine learning models are not trained and tested on the same datasets. They are usually trained on valid authenticated data sets and later used for prediction. This kind of trained data model finds application various domains such as finance industry for volatile market prediction, in the health-care domain for disease identification, anomaly detection in the manufacturing industry and automobile industry, identifying security and privacy violation or intruder detection in the industrial system, etc. For building a good customer base for the AI models, it is necessary that it is trained using trusted datasets and the provenance of these training data to be preserved. Blockchain find a comprehensive solution for capturing the provenance or traceability of the datasets and algorithms used for building the AI models immutably. In this work, we propose a Blockchain based platform for securing e-health data to develop a patient-centric, permission driven data access platform. This work is further extended to build a provenance-based audit trail for the business parties to access the permission given data to run analytics and build trusted AI models in E-health.

A. Contribution of this Article

We have noticed that very few researches have been reported in the Blockchain literature on how to store healthcare data securely and very limited research on patient-driven data sharing and interoperability. To the best of our knowledge, there are no previous works reported that discuss an end-to-end solution on capturing and storing patient information using Blockchain, and also patient-driven data share mechanisms that lead to a provenance-based audit trail for building trusted AI models in Healthcare. In summary, the key contributions made in this article are as follows:

- We propose a framework powered by Blockchain technology that facilitates a patient-driven data sharing mechanism which gives absolute power to the patient to take control over their data.
- We show that a Blockchain powered infrastructure for recording healthcare data including the patient details and medical history lead way for enabling a trusted Artificial Intelligence ecosystem that enhance current centralized AI models.
- We also extend this work for building a provenance-based audit trail for businesses to access the permission on patient-centric data to build trusted AI models and run analytics without disclosing personally identifiable information.

B. Organization of this Article

This article is organized as follows: Section II discusses the preliminaries and background for this study, primarily Blockchain and Electronic Health (e-Health). Section III describes the state-of-the-art of data privacy and security using the Blockchain and also some related works that throw lights on decentralized artificial intelligence which is an emerging area in computing. In section IV, we introduce our proposed frameworks for healthcare data capture, storage, and sharing mechanisms and section V discusses how the proposed frameworks enable a trusted artificial intelligence ecosystem. Section VI concludes the article and our outlook is explained.

II. Preliminaries and Assumptions

A. Blockchain Technology

Blockchain is a distributed ledger technology that provides a secured, immutable, tamper-proofed, distributed data store. Blockchain provides a trustless platform that offers a transparent infrastructure network where a group of non-trusting parties can collaborate and do business without any third party authorization. Instead of centralized record keeping of transaction storage, blockchain offers a decentralized data-store, where every peer in the network keeps a copy of replicated data. The replication, sharing, and synchronization of the data across the peers are managed by the network through some consensus mechanism. The distributed ledger is built on top of a collection of communication protocols that enables decentralized administration of transactions that persisted across geographically separated multiple computing devices. The first application of blockchain is proposed by Satoshi Nakamoto [13], to provide a distributed, transparent technology platform for the much-celebrated cryptocurrency, Bitcoin. The immutability of the data is managed by the consensus mechanism and efficient hashed block linking mechanism. Consensus algorithm makes sure all participating peers have the same state of the network and multiple rounds of endorsements are required for any piece of data to go into a block. A diagram depicting the high level structure of Blockchain is shown in Fig. 1.

The combination of cryptographic hashes, structured hash linking design , multi-party consensus mechanism, and replicated data storage across the nodes provide a trust-full platform by blockchain. A block in Blockchain encapsulates an ordered transaction set, the hash of the previous block, and optionally the nonce of the validating peer [27]. A nonce is nothing but a uniquely identifying piece of information in block creation. In blockchains like Bitcoin or Ethereum, Nonce is the solution to the complex cryptographic puzzle solved by the miner. Every committed block of transactions holds the hash of previous blocks as input. This brings in the immutability of the data records. If any member in the network tampers their copy of shared data, the hash generated from the malicious node will be different from other peers in the network. This causes the falsified node to fail the consensus validation. Unless they are not synced with the rest of the network, they cease to exist in the network. Blockchains are broadly classified into three categories namely public Blockchains, private Blockchains, and consortium Blockchains.

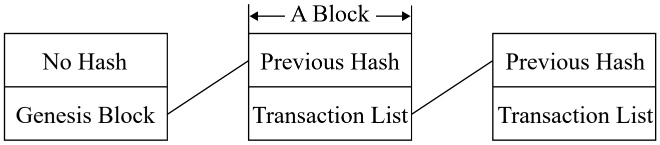


Fig 1. Structural organization of Blockchain.

Public Blockchains are open access blockchain implementations. Membership services are not required for enrolling new users. Any user can be on the blockchain. It is open, which means that any node can access the contents(transactions) of the chain without any restrictions. Any node can take part in the consensus process. The consensus process determines which transactions get added to the Blockchain and maintains the current state of the system. Public blockchains are decentralized meaning no single entity has control over the network. Private Blockchains are privately managed by an individual or organization and only registered members could be part of it. In a private Blockchain, transaction validations are managed by consensus mechanisms or the policy or rules that defines the endorsers or validators. For example, Ethereum Private Network. On the other hand, Consortium blockchains are blockchains managed by a group of institutions who owns a pre-selected number of nodes. These blockchains are partially decentralized [29],e.g.: Corda.

Another classification is permissioned Blockchain and permissionless Blockchain. In permissioned Blockchain, the role of the user can be defined and only the defined user according to the policy has access to the Transaction. For example, Bitcoin and Ethereum are permissionless Blockchains where anyone can be part of the network, the only mandate is he should follow the Proof-of-Work consensus mechanism. However for an enterprise application, to manage the privacy and confidentiality of the data and transaction, industries prefer customizable, access permissions, resource effective Blockchain implementation. This led to the development of permissioned Blockchain [25], [26], [27], [28], [29], [30], e.g.: Hyperledger Fabric, Hyperledger Sawtooth, etc.

B. e-Health and Blockchain

E-health is a term generally used to refer to information and communication technology (ICT) in the healthcare sector which has been evolving since 2000. It has the potential to provide innovative solutions in the medical field encompassing domains like medical informatics, patient healthcare, health care providers, health care professionals, pharmacy, etc. referring to health services, health information, etc. through the Internet and related technologies [33]. As information or data availability becomes seamless through extended Application Programming Interfaces (APIs) and smart devices, patients have a handle to access their medical records and treatment information. Moreover, the data is still getting generated and residing in organizational silos, opening electronic access for the patient to view and manage their data and share it with appropriate stakeholders on demand. The recent trend in healthcare data management is transitioning from organization driven data management towards patient-driven data interoperability to be in-line with GDPR compliance and HIPAA compliance, which demands patients to be the true data owners and not the institutions that generate or create the wealth of data. However, moving towards patient mediated data handling would need to address challenges like implementing new security and privacy protocols for data sharing, authentication, interoperability, confidentiality requirements, electronically verifiable consents, and policy enforcement and governance for data management [9], [1].

C. Securing e-health System and Efficient Data Storage and Sharing

Storage and handling of medical data require a clear definition of access control, authentication, and immutability of data records for ensuring integrity, confidentiality, and accessibility of the health records. Health records encompass both wearable sensor data as well as patient treatment records from healthcare providers. As medical record keeping needs privacy and confidentiality due to data sensitivity, role-based authorization should be implemented for data upload as well as data access. Auditing mechanisms should be enforced to record and monitor the queries and data accesses to be monitored.

Blockchain provides an efficient framework for setting up a private/public network with a secured data access control platform for multi-party business. The potential of Blockchain to create a secured, immutable data sharing platform has been explored previously. Yue et. al. described a methodology of employing a private blockchain for medical record booking and implementing Healthcare Data Gateway (HDG) to enable patients to manage their health records [35]. Ivan [36] proposed a public Blockchain implementation of Personal health record keeping, where medical records are encrypted and stored publicly in a distributed fashion. MedChain [37] employs a permissioned distributed ledger network, where the healthcare stakeholders such as patients, healthcare providers/professionals, pharmacies, etc., could be used to facilitate medication-specific data sharing between agreed entities.

However blockchain technology, has its own share of implications in putting the data records on blockchain. Hence there are few attempts of leveraging blockchain as an enabler for trusted infrastructure in healthcare. They did not assume blockchain as a data layer but, but used smart contracts and blockchain for facilitating management or governance of that data. Zyskind et al. [38] have proposed a blockchain based decentralized access control for managing and handling of encrypted off-chain data. FHIRChain [39], [40] is a healthcare usecase, which is a smart-contract managed system for health data exchange, where the actual clinical data is not residing in the blockchain. But the meta-data of the the clinical records are encrypted and maintained in blockchain as a pointer to actual records in the cloud. Azaria et al. [42] proposed MedRec, which is a prototype built on permissioned blockchain network to enable and manage data sharing and authentication process. MedRec has a novel incentivisation method built around to access the anonymized medical data.

III. Data Privacy and Security using Blockchain: Current State

A distributed ledger management infrastructure implemented for establishing trust in the popular crypto-currency - Bitcoin is now disrupting how businesses function. Trust, being the primary asset of any multi-party involved entities, is achieved by Blockchain by creating a hash-chained record of transactions in an immutable fashion. This is a disruptive technology with many interesting and useful real-world applications that require a trusted data management ecosystem. A cluster of computers (nodes) that decentralize the data management capabilities of an organization, the Blockchain ecosystem, now have the capabilities and potential to move the notion of trust from institutions to mathematically provable infrastructure. This capability exhibited many inherent features that build a trusted platform that captures audit-free transaction data which is essential for any business. This attracted many organizations to implement this technology to enhance their business models and as a result, many variants of Blockchain have emerged with varying levels of success and usability. This section outlines some of the recent advances in using Blockchain technology for better data privacy and security enablement in businesses and other domains.

Very recently, a Blockchain based infrastructure for air traffic management security, authentication, and privacy has been introduced by Ronald J. Reisman [2] which is an open source, permissioned Blockchain based framework. This design enables aircraft privacy and anonymity while providing an efficient and secure mechanism for communicating with air traffic operations support and other stakeholders or entities. This framework provides better security and privacy by enabling higher bandwidth communication channels for private information and deploying smart contracts [2]. The connected and decentralized IoT devices are growing in an exponential fashion which exposes high risk for businesses in terms of privacy and security. The adoption rates for such smart devices are highly dependent on the ability of vendors to provide sufficient sensor data integrity while preserving the privacy of users. A new method for privacy-preserving data certification in IoT networks by leveraging Blockchain technology for protecting sensor data has been introduced recently [3]. The system proposed a blockchain based sensor data protection framework for certifying the data generated and propagated by sensors in an IoT network. The authors claim that the proposed design ensures the tamper-resistant gathering, processing, and exchange of IoT (Internet of Things) sensor data in a privacy-preserving, scalable, and efficient manner [3].

A privacy-friendly platform for healthcare data in cloud-based on Blockchain environment [4] is introduced in the literature that presented a patient-centric healthcare data management system powered by Blockchain technology for attaining privacy. This work used cryptographic functions for encrypting patient's data for ensuring pseudo-anonymity [4]. A novel framework for addressing security and privacy issues in healthcare using modified Blockchain models which are apt for IoT devices was introduced by Ashutosh Dhar Dwivedi et. al. [5]. The additional security and privacy features of their modified Blockchain model are based on basic cryptographic properties that make IoT based healthcare networks more secure and anonymous [5]. A Blockchain based learning healthcare system that could foster the willingness of a patient to contribute to research and learning by providing sufficient control over health data is proposed by Marielle S. Gross and Robert C. Miller [6]. This system also enables mechanisms for meta-analysis without even exposing individual details of the patient and this will allow protecting the privacy of patients [6].

Another system for medical data management on Blockchain for preserving privacy, that uses Hyperledger Fabric to store encrypted data has been proposed by Haibo Tian, Jiejie He and Yong Ding [7]. The authors proposed the idea of a shared key that could be reconstructed by legitimate parties before starting the treatment of patients. The approach uses sibling interactable function families to establish the shared key for data encryption [7]. Design of a storage scheme for managing and sharing medical records is proposed by Yi Chen et. al. The storage and sharing schemes employed by the proposed system do not depend on any third-party and there is no single entity that has absolute power to make the process affected [8].

William J. Gordon and Christian Catalini looked at how Blockchain technology can facilitate patient-driven interoperability by using features such as digital access rules, data aggregation, data liquidity, patient identity, and data immutability [9]. The authors concluded that patient-driven interoperability is a trendsetter in the healthcare industry having many challenges, Blockchain may facilitate the transition from a healthcare provider centric ecosystem to a patient-centric ecosystem [9]. A new framework called ModelChain was introduced to adapt Blockchain technology for privacy-preserving machine learning systems [10]. In this system, the machine learning model parameter estimation is done without disclosing any patient health information and the proposed framework applies Blockchain technology to solve the privacy-preserving predictive modeling task for healthcare that will facilitate and increase the potential of interoperability between various stakeholders [10]. A new data sharing solution built on Blockchain that addresses two prevalent issues with the healthcare domain such as protecting sensitive health information and deployment and installation of Blockchain software across diverse hospital networks has been proposed [11]. The newly proposed innovative architecture addressed critical data security, deployment, and installation challenges and provides the healthcare community with an approach to connect diverse and heterogeneous providers while protecting sensitive healthcare data [11].

A. Trend Towards Trusted and Decentralized Intelligence

Advances in computing and communication technologies and heavy investments in exponential technologies such as Artificial Intelligence caused hundreds of millions of devices to run AI / ML models. But these devices and AI models are built and run by some centralized mega-corporations that makes the entry difficult for other small entities into this ecosystem. The organizations who collect and curate data may not be running the AI models most of the time and the end users will be mostly another set of people. This situation demands Gabriel Axel Montes and Ben Goertzele in their work on Distributed, decentralized, and democratized artificial intelligence [14] claims that decentralizing the AIwill generate more equitable development of Artificial Intelligence and Artificial General Intelligence. The authors also state that the progress towards decentralized AI will also create the infrastructure for coordinated action between AIs that will significantly facilitate the evolution of AI into true Artificial General Intelligence that is both highly capable and beneficial for humanity [14]. A recent study on the decentralized marketplace data using Blockchain technology which aims to provide confidence among multiple stakeholders in the system. This specific study on the supply chain domain ensures four different levels of data management such as data provision, data delivery, rights management, and producer internal sources [33].

In this work, we propose a Blockchain powered framework for securely storing patient sensitive information and other associated healthcare data such as treatment history, by enabling a patient-driven data sharing mechanism. The core idea of the proposed work is to make the patient the true owner of data and thereby giving absolute power to decide which data needs to be shared with the healthcare providers. The audit trail of the complete transaction details will be recorded in a tamper-proof digital ledger which is an auditable record for reference. The digital identity management of the proposed framework will generate a unique identity for a patient to which the treatment history and other medical records can be appended and this makes a moving electronic health record management ecosystem which is patient-managed. In addition to this, we propose how a Blockchain based data management framework that employs audit trail features can enable the notion of trust on the data captured among multiple entities to make use of a trusted but decentralized AI model.

IV. Blockchain for Securing Healthcare Data

Recent researches on data management in healthcare shows a clear transition from an organization (healthcare provider) centered data management to patient-driven data management mechanisms. Even though the idea of transforming patients as the true owners of their data is not new, the very recent regulations such as the European Union General Data Protection Regulation (GDPR) fueled this transformation. According to European Union, the GDPR (https://eugdpr.org/), is the most important and revolutionary change in data privacy regulation in the last twenty years which will reshape how data is handed across every sector - be it healthcare to banking and beyond. While many other sectors will change how they manage data, everybody is eyeing the healthcare domain and is very keen to analyze how the regulations mentioned above are going to be implemented in healthcare. We can see that regulations in healthcare such as the 2009 Health Information Technology for Economic and Clinical Health Act (HITECH) introduced in the United States has increased the digitization of Electronic Health Records (EHR), the sharing of such records among different hospitals and other healthcare providers are lagging much behind. This is due to the operational, technical and most importantly the privacy-related issues prevalent specifically in the healthcare domain [9].

Interoperability in healthcare is often interpreted as the ability to exchange medical records electronically between hospitals managed by the same business unit or between multiple individual hospitals, to some extent. Here, the data management, protection, and exchange is the responsibility of the business units such as hospitals and the patient normally does not come into the picture. In such a scenario, the patients have no clue where their data is being stored or with whom their data has been shared. Recently, there is a paradigm shift from the traditional business-centric approaches to a patient-driven or patient-centric interoperability where the patient becomes the absolute owner of their data. However, this shift introduces an array of new challenges in terms of technology, security, privacy and governance and still many of the issues related to these aspects have not been solved for even business-centered interoperability use-cases. To accelerate the facelifting of patient-centered data management, we need new technology supported initiatives and systems that not only solve the issues associated with the traditional interoperability but also pave ways for patient-centered interoperability easily.

Blockchain, the technology infrastructure behind Bitcoin[13], is a novel technology that has built-in features such as tamper-proof, audit-ability, decentralization, cryptographic signatures and hashing, etc. makes it appropriate for storing data securely which leads to improved interoperability[12][13]. Blockchain technology has the potential to facilitate and accelerate the transition from an institution-centered data management ecosystem to patient-centered or patient-controlled data management by incorporating data aggregation, data access rules, immutability, and patient identity management[9]. Identifying and incorporating the benefits provided by Blockchain technology in facilitating privacy preserving and patient-centered data management, we have attempted this problem in detail and proposed new frameworks for patient data management and sharing. We have also extended these frameworks for building a provenance based Blockchain model for an audit trail mechanism that leads towards a trusted artificial intelligence ecosystem for healthcare. A list of blockchain features and their application in patient-driven data sharing mechanisms are given in Table I.

TABLE I. Features of Blockchain and Application in Patient Driven Data Management and Data Sharing

feature	APPLICATION
Immutability	Patient information, treatment information and other healthcare records can be securely shared among multiple providers lowering the risk of attacks and loss and thus avoiding the need for audit-ability and verifiability.
Identity Management	Every patient can have a unique digital identifier which can be stored efficiently on Identity Management Blockchain and all the EHR and other treatment information can be mapped against this identity without the worry of being tampered
Decentralization	Since there is no notion of centrality, the clinical Decentralization and other treatment data can be accessed among multiple hospitals and near realtime records can be accessed.
Access Rules	Patients can give consent for accessing their Access Rules medical records through self-executing codes on blockchain called “smart contracts” and which in turn can keep audit trail of data access transactions

A. Blockchain based Data Privacy and Security Systems

Here, we propose a Blockchain based digital identity and consent management mechanism for healthcare which is shown in Fig.2.

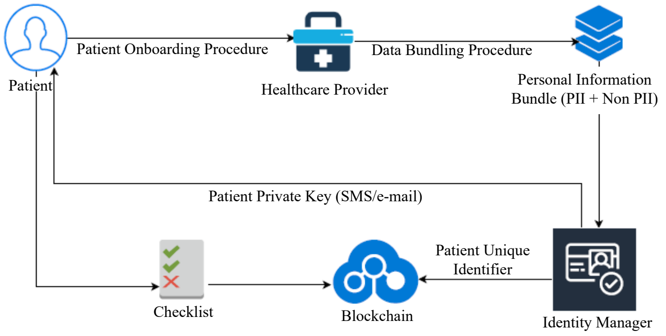

Patient
Patient Onboarding Procedure
Healthcare Provider
Data Bundling Procedure
Personal Information Bundle (PII + Non PII)
Patient Private Key (SMS/e-mail)
Patient Unique Identifier
Checklist
Blockchain
Identity Manager

Fig. 2. Proposed Patient Registration / On-boarding Process.

- **Patient Registration / On-boarding Procedure:** The healthcare provider (HCP) collects the complete information including personally identifiable information (PII) and non-personally identifiable information (non-PII) of the patient through the patient on-boarding procedure. The HCP then segregates PII and Non-PII elements from the data and creates a data bundle (PII + Non-PII) and passes this information into a Blockchain based identity manager. If the hospital or healthcare provider already has a database based EHR management system, then the Non- PII data can be pushed to the same system. The personally identifiable information will be captured in an offline database and the hash of this record will be stored in the Blockchain. In order to link the offline record with the Blockchain transactions, we employ a foreign key generated using a hash function. This allows the patient to update the personally identifiable information seamlessly. However, we only describe a complete blockchain based system assuming that the hospital does not already have a technology infrastructure for dealing with patient data and need a complete tamper-proof mechanism to deal with patient information and other electronic health records. The identity manager (IM) module of the proposed framework generates a unique identifier for the on-boarded patient and record the same in the Blockchain. Simultaneously, the IM will generate a private key for the patient and communicate the same through an SMS or e-mail to the registered mobile phone number or e-mail address of the patient. During onboarding or registration, the healthcare provider gives a checklist of data that needs to be captured and/or shared and the patient needs to check/un-check the list of items according to their willingness to get the data captured by the provider. On submitting the final checklist, the same will be updated in the Blockchain against the private key of the patient. Finally, a patient will have a unique identifier and an associated checklist of items related to their data sharing consents. Such a system would facilitate a patient-centered and patient-driven data sharing ecosystem where the patient has absolute power on their data and only with his consent, the data can be viewed or shared among various healthcare providers. Revenue models can be built around data sharing platforms. Based on the permission agreement

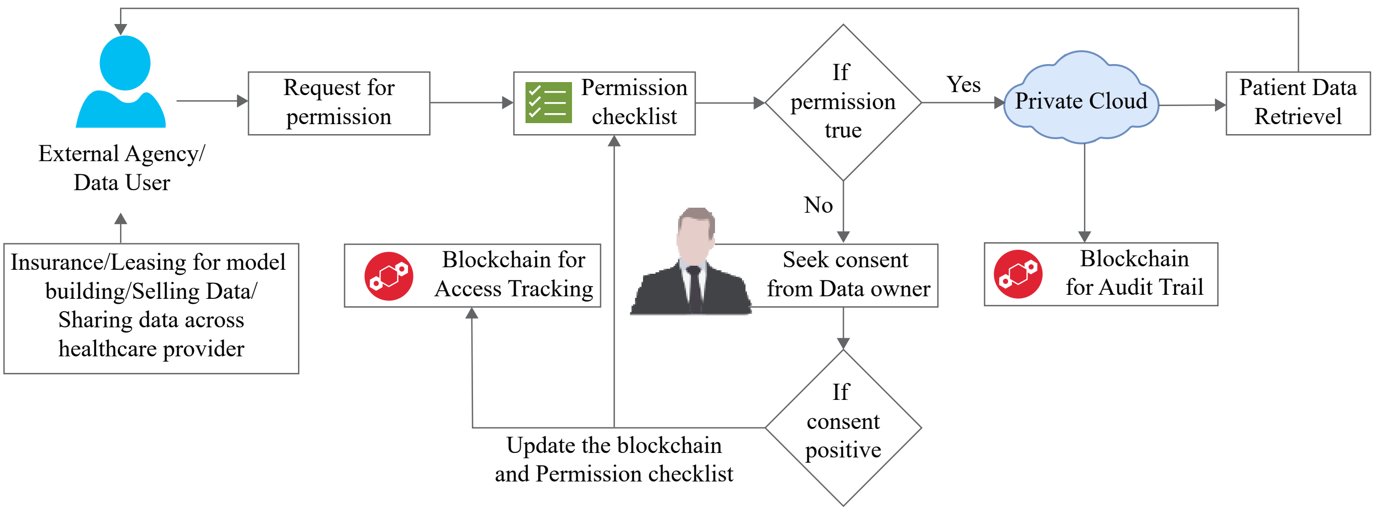

External Agency/
Data User
Request for permission
Permission checklist
If permission true
Yes
No
Private Cloud
Patient Data Retrievel
Insurance/Leasing for model building/Selling Data/ Sharing data across healthcare provider
Blockchain for Access Tracking
Seek consent from Data owner
Blockchain for Audit Trail
If consent positive
Update the blockchain and Permission checklist

Fig. 3. Patient Data Request Process.

data could be shared across various stakeholders under stipulated pre-published pricing policies. Blockchain offers excellent platform for immutable record keeping of the rule based condition of data sharing policy and tamper-proof distributed compute-storage, which provides a secure environment for executing multi-party business. Self-executing smart contracts mandates the policy enforcement and the audit trial logging mechanism in blockchain offers the traceability of all executed transactions and smart contracts.

- **Data Request Procedure:** In this proposed system, patient health records have been captured and stored in the Blockchain with an appropriate checklist of permissions. When a healthcare provider or other users request for this recorded data, they need consent from the owner of the data. To enable this, a request for granting the permission will be submitted to the network with a list of required data items. The requested data records will be cross-checked with the patient published checklist. Consented data items for which the patient has given consent will be shared with the requester. Simultaneously, the audit trail of the data access along with complete request information will be stored in a Blockchain exclusively for audit trail. The data owner will also be alerted regarding data access. In case of data access permission revocation, a change in the permission will be recorded in the audit trail Blockchain as a new transaction and an intimation will be sent to the data requester. The proposed data request for healthcare providers is shown in Fig.4.

In case the data requester/user is requesting data items which have no access rights/consent given by the patient, then this request will be sent to the patient asking for his consent. It is the sole discretion of the patient to grant access to the data items. If the patient gives a new consent for the requested item, then the same will be first updated in the permission checklist and also on the Blockchain for access tracking. In this way, the patient can grant or deny access to data items that are owned by him.

B. Towards a Trusted Decentralized Artificial Intelligence Ecosystem using Blockchain

Heavy investments in Artificial Intelligence and Machine Learning caused trained ML models to become more accessible by various stakeholders which in turn benefits the end users directly and indirectly. In such a decentralized intelligent ecosystem, the various participants need to trust the model which requires provenance information on how the data has been captured and trained. According to Sarpatwaret. al. [41] the provenance data model for an AI/ML model includes the following information:

- Complete details of the data used to build the AI model.
- Details of the model pipeline.
- Information on the training process.
- Details of the updates to the trained model.
- Testing details of the trained model.

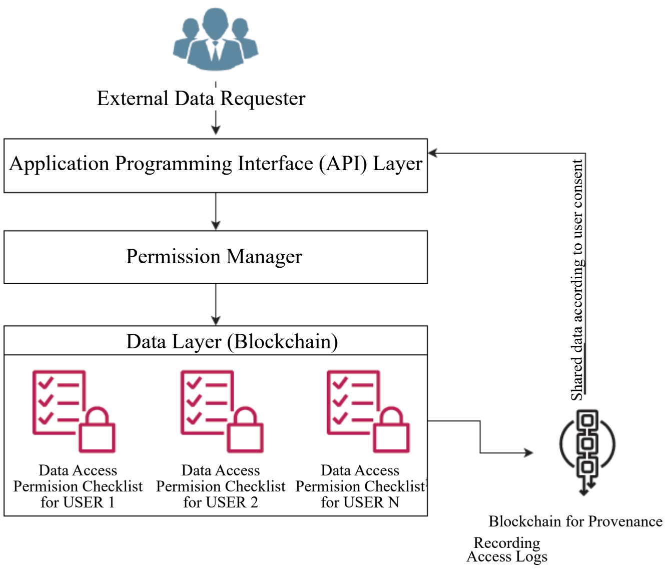

External Data Requester
Application Programming Interface (API) Layer
Permission Manager
Data Layer (Blockchain)
Data Access Permision Checklist for USER 1
Data Access Permision Checklist for USER 2
Data Access Permision Checklist for USER N
Shared data according to user consent
Blockchain for Provenance
Recording Access Logs

Fig. 4. Data Request Process for Healthcare Providers.

This information needs to be stored in a tamper-proof, decentralized and immutable manner to harness the trust of the decentralized AI model. Blockchain is an ideal solution to tackle this situation which can capture the complete information about the trained model right from the data collection to the model validation to testing. This facilitates and creates a transparent and trusted decentralized artificial intelligence ecosystem where various participants share the credible data or trained model. Finally, there won't be a trusted centralized organization that manages the AI model generation; rather there will be decentralized and trusted participants that can contribute to the overall training and evaluation of AI models. The healthcare domain is not an exception, thus we need a Blockchain based provenance mechanism for capturing patient data right from onboarding (with appropriate consent from the data owner) which is powered by Blockchain. We propose a framework for provenance (Refer Fig. 3) that aims to provide or give access to captured healthcare data for external vendors to build Artificial Intelligence / Machine Learning models or run analytics. This is done by exposing REST APIs with appropriate authentication and permission management procedures. The data layer stores and manages data access permission checklist for every individual patient onboard in the system. On accessing data after getting sufficient consent on the data items requested, complete access logs will be recorded in the provenance Blockchain which makes the data credible for other vendors to build AI/ML models which will be beneficial for many number of services and applications.

C. Experimental Setup

Here, the implementation details of our proposed frameworks are discussed. We have created an experimental testbed using Hyperledger Sawtooth which is an enterprise Blockchain platform for creating and managing permissioned ledgers. We have used the native installation using Ubuntu 16.04 with Sawtooth version 1.0 which is available at https://github.com/hyperledger/sawtooth-core/ for simulating a private Blockchain network locally. Since we need to implement activities of different stakeholders such as healthcare providers, patients, and regulators, a three node network with three validators have been configured in the network. Also, Sawtooth validators, REST APIs and transaction processors are created and initiated in the already created network. We have used Python SDK for Sawtooth available at https: //github.com/hyperledger/sawtooth-sdk-python. For comparing the operational advantages of a database based system w.r.t a blockchain driven health- care data managed system, we setup a MySQL database, where the DB provide the single, centralized instance of data source. However this DB based system is not free from data-manipulation or data forging attempts by malicious agents.

In a blockchain driven system, data could be either stored off-chain and the meta-data of the stored information could be immutably secured in blockchain. However in this work we loaded the data directly into blockchain and access permissions to the data are managed by smart contracts. The subtle advantage of the blockchain driven system over database only based system is the immutability of the data records in blockchain, be it permission settings permitted by the patient, data access audit trial, data logging into the system, etc. Revenue models built around it would need a transparent eco-system

D. Dataset Description

The main challenge faced by researchers in the healthcare domain is the availability of public datasets. Since healthcare records are very sensitive, they are mostly not available publicly. However, there are very few anonymized datasets accessible online. The proposed work requires data records with patient identifiers such as name, date of birth and other contact details, the anonymized datasets will not suit our needs. For this work, we have used a hybrid method for collecting anonymized data records and appended dummy data records with personally identifiable information. The MIMIC-III dataset is a freely available healthcare dataset containing deidentified health-related data associated with over 40000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012 [32][33]. For the current experiment, we For the current experiment, we have only used 1000 records due to scalability and other data management constraints.

E. Implementation Details

In our proposed framework, the patients need not run a network node and they are treated as clients. Infrastructure will be provide REST APIs for accessing and updating their personal information, data access permissions and other consents. We have implemented a web application that captures complete data of patients including PII and non-PII while on-boarding at a health-care provider. For this experiment, we have on-boarded 1000 patient records taken from MIMIC-III dataset (after appending identifiable information) and has recorded the same in the Blockchain (Refer Fig. 1) along with default consents for all records. In order to compare the operation of a simple database system with respect to a Blockchain mediated system, we have implemented a standalone MySQL database (DB) for record keeping. MySQL database holds patients personal details in patient-personal table and treatment details in imaging table, diagnosis table and pharmacy table. Access permissions are stored in *usr-access-request* and *usr-access-perm* tables. Data logging regarding the permissions set by the user and details of the data access by the user are persisted in audit data table.

Our data request framework (Fig. 3) requires two separate Blockchains: One will store patient information and data access consents and another one for capturing the audit trail of the queries/ requests made by any requester. We tried accessing information of different patients and found that the data requester (client) can only access those data items from the patient record for which the patient has given access to. In addition to this, we have also tested whether access logs are added to the Blockchain for audit trail. We later updated the default data access permission of patients and repeated the data request process. This is to ensure that the changes are updated in the system while updates are made into the system through the checklist API provided.

Finally, we have set-up provenance Blockchain using Hyperledger Sawtooth for implementing our framework for external data request by third parties. Whenever a data request has been initiated by external data requester, the permission manager checks and verify whether the requester has sufficient privileges to use the REST APIs for data access. To check this functionality, we have added another client to the provenance Blockchain network and given access to the APIs. When a data access request is received, the individual data access permission (patient consent checklist) is referenced and on granting data records, complete access logs including the requester details and granular level data details will be recorded in the Provenance Blockchain. This ledger keeps track of all data access information which will act as a credible source of data on which decentralized artificial intelligence or machine learning models can be built and share.

In-order to compare the operation of a blockchain based system with other legacy system, we assume, a malicious actor from the healthcare provider is trying to tamper with the data. This is achieved by manipulating the access permissions set by the patient and attempts to alter the pricing strategy to benefit the organisation. However as blockchain maintains immutable and tamper proof logging and record keeping of the transactions in distributed manner, any malicious attempts in local data manipulation, would be caught during consensus evaluation. Also, every attempt to login into the system is also captured in the audit module, details of all unauthorized access to the patient data can be analyzed in future.

V. Discussions and Lessons Learnt

The proposed approach was implemented in a controlled environment with a restricted number of user profiles. Blockchain technology for enterprise applications are still in an infancy stage. We faced issues on integrating various aspects of the proposed system in terms of scalability, usability, accessibility and interoperability. The key challenges identified during our pilot implementation of the proposed method are outlined below:

- Interoperability between cloud, legacy system, mobile application and Blockchain.

- Organizational policies and restrictions on storing data over cloud.
- The inhibition from patients to store and acquire data by the cloud services and the associated cost for the same.
- Absence of data protection regulations and compliances from the regulatory bodies in healthcare.

Comparison Note: Current centralized systems use dedicated database servers deployed in organization premises or cloud-based services which brings the notional of centrality to a greater extent. Maintaining a single version of the truth and deriving actionable insights from data accumulated at local repositories is always a complex task. The extraction, transformation, and loading (ETL) process of data from local repositories are complex and thus a major hindrance to the quick decision-making process. On the other hand, a Blockchain always has a single version of the truth that is maintained in a decentralized manner. As every node in the network has the same copy of ledger which is audit-free, the near real-time analysis can be performed. While applications that use database backend show better performance and scalability, the major bottleneck of Blockchain is found to be scalability, interoperability and the delay in committing transactions. The technology is still in an infancy stage and requires a lot of improvements in terms of scalability and interoperability that may enable higher transaction throughput and better-managed services.

VI. Conclusions and Future Work

This work proposes a platform that enables patients to be the true owner of their data. The frameworks support secure, immutable, auditable and traceable platform powered by Blockchain that enables patients to manage their healthcare data. We show that our proposed frameworks can take away the notion of trust from centralized organizational infrastructure to decentralized mechanisms. The system also caters to the requirements for building a trusted AI ecosystem with provenance. The Proof-of-Concept (PoC) implementation of our proposed frameworks in a Permissioned Blockchain shows that there is significant potential to explore further. This PoC with limited data management capabilities needs to be scaled-up to incorporate new stakeholders and other entities. Patient mobility and easiness of managing permissions are issues that need to be addressed and the authors will be working on these dimensions in the future.

Acknowledgment

This work is supported by the Back to Lab Programme research fellowship (Ref: Order No.1281/2016/KSCSTE) from Women Scientists Division (WSD), Kerala State Council for Science, Technology and Environment (KSCSTE).The authors would like to thank all the researchers, engineers and other staff members of Kerala Blockchain Academy and Data Engineering Lab at Indian Institute of Information Technology and Management-Kerala, for their valuable comments that significantly improved the quality of this paper. The authors also acknowledge the staff members for providing infrastructure facilities for conducting an experiment of this scale.

References

[1] T. Mcghin, K.-K. R. Choo, C. Z. Liu, and D. He, “Blockchain in healthcare applications: Research challenges and opportunities,” *Journal of Network and Computer Applications*, vol. 135, pp. 62–75, 2019.

[2] R. Reisman, “Blockchain Serverless Public/Private Key Infrastructure for ADS-B Security, Authentication, and Privacy,” *AIAA Scitech 2019 Forum*, 2019.

[3] M. Chanson, A.Bilgeri, D., Fleisch, E., and Wortmann, F. “Privacy-Preserving Data Certification in the Internet of Things: Leveraging Blockchain Technology to Protect Sensor Data,” *Journal of the Association for Information Systems,* 2019.

[4] A. A. Omar, M. Z. A. Bhuiyan, A. Basu, S. Kiyomoto, and M. S. Rahman, “Privacy-friendly platform for healthcare data in cloud based on blockchain environment,” *Future Generation Computer Systems*, vol. 95, pp. 511–521, 2019.

[5] A. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, “A Decentralized Privacy-Preserving Healthcare Blockchain for IoT,” *Sensors*, vol. 19, no. 2, p. 326, 2019.

[6] M. S. Gross and R. C. Miller, “Ethical Implementation of the Learning Healthcare System with Blockchain Technology,” *Blockchain in Healthcare Today*, vol. 2, 2019.

[7] H. Tian, J. He, and Y. Ding, “Medical Data Management on Blockchain with Privacy,” *Journal of medical systems*, vol. 43, no. 2, article 26, 2019.

[8] Y. Chen, S. Ding, Z. Xu, H. Zheng, and S. Yang, “Blockchain-Based Medical Records Secure Storage and Medical Service Framework,” *Journal of Medical Systems*, vol. 43, no. 1, 2018.

[9] W. J. Gordon and C. Catalini, “Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability,” *Computational and Structural Biotechnology Journal*, vol. 16, pp. 224–230, 2018.

[10] T. T. Kuo, and L. Ohno-Machado (2018), “Modelchain: Decentralized privacy-preserving healthcare predictive modeling framework on private blockchain networks,” arXiv preprint arXiv:1802.01746.

[11] M. A. Cyran, “Blockchain as a Foundation for Sharing Healthcare Data,” *Blockchain in Healthcare Today*, 2018.

[12] C. Catalini and J. Gans, “Some Simple Economics of the Blockchain,” 2016.

[13] S. Nakamoto“Bitcoin: A peer-to-peer electronic cash system.”, 2008

[14] G. A. Montes and B. Goertzel, “Distributed, decentralized, and democratized artificial intelligence,” *Technological Forecasting and Social Change*, vol. 141, pp. 354–358, 2019.

[15] What Is Patient Engagement? | Evariant: The Leading Healthcare CRM Solution. [online] Available at: <https://www.evariant.com/faq/why-is-healthcare-data-management-important> [Accessed 23 June 2020].

[16] H. Löhr, A.-R. Sadeghi, and M. Winandy, “Securing the e-health cloud,” *Proceedings of the ACM international conference on Health informatics - IHI 10*, 2010.

[17] D. Kotz, C. A. Gunter, S. Kumar, and J. P. Weiner, “Privacy and Security in Mobile Health: A Research Agenda,” *Computer*, vol. 49, no. 6, pp. 22–30, 2016.

[18] M.A.Sahi, H.Abbas, K.Saleem, X.Yang, A.Derhab, M.A.Orgun, W. Iqbal, I. Rashid, and A. Yaseen, “ Privacy preservation in e-healthcare environments: State of the art and future directions,” *IEEE Access*, vol. 6, pp.464-478. 2017.

[19] M. Panner, “Blockchain In Healthcare: How It Could Make Digital Healthcare Safer And More Innovative,” 2019, <https://www.forbes.com/sites/forbestechcouncil/2019/06/18/blockchain-in-healthcare-how-it-could-make-digital-healthcare-safer-and-more-innovative/#2ff4bacf3e5a> [Accessed 23 June 2020].

[20] W. J. Gordon and C. Catalini, “Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability,” *Computational and Structural Biotechnology Journal*, vol. 16, pp. 224–230, 2018.

[21] W.J. Gordon, A. Wright and A. Landman. Blockchain Technology in Health Care: Decoding the hype. NEJM Catal2017 https://catalyst.nejm.org/decoding-blockchaintechnology-health/, Accessed date: 28 March 2018.

[22] A. J. Holmgren, V. Patel, and J. Adler-Milstein, “Progress In Interoperability: Measuring US Hospitals’ Engagement In Sharing Patient Data,” *Health Affairs*, vol. 36, no. 10, pp. 1820–1827, 2017.

[23] Q. Nasir, I.A. Qasse, M. Abu Talib, and A.B Nassif, “Performance analysis of hyperledger fabric platforms,” *Security and Communication Networks*, 2018.

[24] A. Brando,H. So Mamede,and R. Gonalves (2019, April), “Trusted Dataset Marketplace,” In *World Conference on Information Systems and Technologies* (pp. 515-527). Springer, Cham.

[25] C. Cachin (2016, July), “Architecture of the hyperledger blockchain fabric,” In *Workshop on distributed cryptocurrencies and consensus ledgers* (Vol. 310, No. 4).

[26] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. D. Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich, S. Muralidharan, C. Murthy, B. Nguyen, M. Sethi, G. Singh, K. Smith, A. Sorniotti, C. Stathakopoulou, M. Vukolić, S. W. Cocco, and J. Yellick, "Hyperledger fabric," *Proceedings of the Thirteenth EuroSys Conference*, 2018.

[27] V. Dhillon, D. Metcalf, and M. Hooper, "The Hyperledger Project," *Blockchain Enabled Applications*, pp. 139–149, 2017.M. Swan (2015). *Blockchain: Blueprint for a new economy*. " O'Reilly Media, Inc."

[28] Gorenflo, Christian, Stephen Lee, Lukasz Golab, and Srinivasan Keshav. "Fastfabric: Scaling hyperledger fabric to 20,000 transactions per second." In 2019 *IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, pp. 455-463. IEEE, 2019.

[29] Benhamouda, F., Halevi, S. and Halevi, T., 2019. Supporting private data on hyperledger fabric with secure multiparty computation. *IBM Journal of Research and Development*, *63*(2/3), pp. 3:1-3:8.

[30] K.Olson, M.Bowman, J. Mitchell, S. Amundson, D. Middleton, and C. Montgomery, "Sawtooth: An Introduction. The Linux Foundation," 2018.

[31] I. C. Lin and T. C. Liao, "A survey of blockchain security issues and challenges," *IJ Network Security*, vol. 19, no. 5, 653-659. 2017.

[32] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, article 160035, 2016.

[33] Johnson, A., Pollard, T., & Mark, R. (2019). MIMIC-III Clinical Database Demo (version 1.4). *PhysioNet*. https://doi.org/10.13026/C2HM2Q.

[34] H. Oh, A. Jadad, C. Rizo, M. Enkin, J. Powell, and C. Pagliari, "What Is eHealth (3): A Systematic Review of Published Definitions," *Journal of Medical Internet Research*, vol. 7, no. 1, 2005.

[35] X. Yue, H. Wang, D. Jin, M. Li, and W. Jiang, "Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control," Journal of Medical Systems, vol. 40, no. 10, 2016.

[36] D. Ivan. "Moving toward a blockchain-based method for the secure storage of patient records", *NIST/ONC*; 2016.

[37] WJ. Gordon, A. Landman, "Secure, decentralized, interoperable medication reconciliation using the Blockchain",*NIST/ONC*, 2016.

[38] G. Zyskind, O. Nathan, and A. sandy Pentland, "Decentralizing Privacy: Using Blockchain to Protect Personal Data," *2015 IEEE Security and Privacy Workshops*, 2015.

[39] "Health Level Seven International," *Health Level Seven International - Homepage*. [Online]. Available: http://www.hl7.org/. [Accessed: 22-Jun-2020].

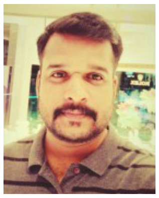
[40] P. Zhang, J. White, D. C. Schmidt, G. Lenz, and S. T. Rosenbloom, "FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data," *Computational and Structural Biotechnology Journal*, vol. 16, pp. 267–278, 2018.

[41] K. Sarpatwar, R. Vaculin, H. Min, G. Su, T. Heath, G. Ganapavarapu, and D. Dillenberger, "Towards Enabling Trusted Artificial Intelligence via Blockchain," *Policy-Based Autonomic Data Governance Lecture Notes in Computer Science*, pp. 137–153, 2019.

[42] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, 2016, August. "Medrec: Using blockchain for medical data access and permission management," In *2016 2nd International Conference on Open and Big Data (OBD)* (pp. 25-30). IEEE.

Jennath H.S.

Jennath received a B.Tech in Electronics and Communication Engineering from University of Calicut in 2005. She worked as Technology Lead in IT companies like Infosys and IBS Software Services after her graduation. She did her M.Tech in Wireless Networks and Application from Amrita Vishwa Vidyapeetham in 2015. She is currently pursuing her Ph.D. from Cochin University of Science and Technology and her research centre is Indian Institute of Information Technology and Management – Kerala (IIITM-K). Her research interests include Vehicular Communication, Wireless Networks, Machine Learning and Blockchain.



Anoop V.S.

Anoop has completed his research leading to Ph.D. in Computer Science with specialization in Artificial Intelligence, from Indian Institute of Information Technology and Management – Kerala (IIITM-K) under Cochin University of Science and Technology. He has also completed a Master of Philosophy (M.Phil.) with specialization in Artificial Intelligence from IIITM-K and Master of Computer Applications from IGNOU, New Delhi. Anoop is currently associated with Kerala Blockchain Academy in the role of a Senior Scientist and Head of Research &Training. Prior joining KBA, Anoop has worked with Etisalat and Ethihad Airways as part of Dubai Future Accelerator and Fika Labs Program respectively in the United Arab Emirates. He is an experienced Software Engineer with over six years experience and also served IGNOU as a faculty member for its School of Computer and Information Sciences. Anoop's primary research interests include Applied Artificial Intelligence using Text Mining, Natural Language Processing (NLP) and Information Retrieval and Blockchain. He has several publications in his credit including international and national journals, book chapters and conference proceedings in major venues of Artificial Intelligence and Knowledge Management.

Asharaf S.

Dr. Asharaf S is a Professor at Indian Institute of Information Technology and Management – Kerala. He is also serving as a Professor-in-Charge for Kerala Blockchain Academy (KBA) and also a visiting faculty in Indian Institute of Space Science and Technology, Trivandrum and as a Mentor in Kerala Startup Mission. He received his Ph.D. and Master of Engineering degrees in Computer Science from Indian Institute of Science, Bangalore. He graduated in Computer Engineering from Cochin University of Science and Technology. After his PhD he has worked with America Online (AOL) and IIM Kozhikode. He is a recipient of IBM outstanding PhD student award 2006, IBM Shared University Research Grant, 2015 and IBM Open Science Collaboration Programme grant, 2017. He has published three books and more than 30 research papers in international journals and conferences. His areas of interest include technologies and business models related to data engineering, machine learning, information retrieval and blockchains.