Peer-to-Peer Networking and Applications (2021) 14:3012–3028
https://doi.org/10.1007/s12083-021-01153-y

HonestChain: Consortium blockchain for protected data sharing in health information systems

Soumya Purohit[1] · Prasad Calyam[1] · Mauro Lemus Alarcon[1] · Naga Ramya Bhamidipati[1] · Abu Mosa[1] · Khaled Salah[1,2]

Received: 25 November 2020 / Accepted: 2 April 2021 / Published online: 3 May 2021

Abstract
Healthcare innovations are increasingly becoming reliant on high variety and standards-compliant (e.g., HIPAA, common data model) distributed data sets that enable predictive analytics. Consequently, health information systems need to be developed using cooperation and distributed trust principles to allow protected data sharing between multiple domains or entities (e.g., health data service providers, hospitals and research labs). In this paper, we present a novel health information sharing system viz., HonestChain that uses Blockchain technology to allow organizations to have incentive-based and trustworthy cooperation to either access or provide protected healthcare records. More specifically, we use a consortium Blockchain approach coupled with chatbot guided interfaces that allow data requesters to: (a) comply with data access standards, and (b) allow them to gain reputation in a consortium. We also propose a reputation scheme for creation and sustenance of the consortium with peers using Requester Reputation and Provider Reputation metrics. We evaluate HonestChain using Hyperledger Composer in a realistic simulation testbed on a public cloud infrastructure. Our results show that our HonestChain performs better than the state-of-the-art requester reputation schemes for data request handling, while choosing the most appropriate provider peers. We particularly show that HonestChain achieves a better tradeoff in metrics such as service time and request resubmission rate. Additionally, we also demonstrate the scalability of our consortium platform in terms of the Blockchain transaction times.

Keywords Blockchain · Health information sharing · Hyperledger · Reputation · Trust

1 Introduction

The increase in data-driven methods to create healthcare innovations requires use of voluminous, high variety and standards compliant (e.g., HIPAA [1], common data model [2]) distributed datasets that enable predictive analytics. These data sets have to be shared by data custodians or providers from multiple domains (e.g., health data service providers, hospitals and research labs) with data requesters (e.g., re- searchers, clinicians) with stringent processes that ensure information assurance and auditability. Handling health big data has many challenges and requires innovative technolo- gies to ensure efficient and secure data processing. Authors in [3, 4] provide comprehensive surveys on big data handling challenges. Security aspects are considered in [5] for handling health big data. Health information systems are being developed to allow such protected data sharing from/− between multiple domains [6]. All of them involve autho- rized approval for receiving requested data as part of a data governance process, which also handles interoperability of heterogeneous data sets for the staging data [7], and subse- quent query/storage of the requested data for analysis.

The authorization step in health information systems can cause data access bottlenecks due to trust issues among the data custodians and requesters. Such trust issues lead to the fear of "Loss of Value" for the data being provided by the data custodians. To cope with risks associated with "Loss of Value" and to ensure assurance/auditability in data ac- cess [8], data custodians use high-touch methods that require a governance committee to *manually* approve data requests. Consequently, manual approvals in the data access transactions cause long queues of data requests. In addition, hu-

This article is part of the Topical Collection: *Special Issue on Blockchain for Peer-to-Peer Computing*
Guest Editors: Keping Yu, Chunming Rong, Yang Cao, and Wenjuan Li

✉ Khaled Salah
calyamp@missouri.edu

[1] University of Missouri-Columbia, Columbia, MO, USA

[2] Khalifa University, Abu Dhabi, UAE

man error in the forms filled out by data requesters can prolong the deliberation of the governance committee and cause over-provisioning or under-provisioning of data queries for data analytics/visualization. These factors cause frictions in the innovation process and causes delayed patient care decisions, which ultimately leads to a "Loss of Opportunity" [9] in the requested data.

Figure 1 illustrates four cases involving either authorized access or access with Loss of Opportunity/Value. In the first case, Requester 1 submits a data request that does not com- ply with the security standards, and as a result data is not provided and a Loss of Opportunity is experienced. In the second case, Requester 2 submits a data request that is han- dled without the proper auditability mechanisms, which po- tentially leads to Loss of Value. In the third case, Requester 3 submits a request that complies with the security standards, and hence data is provided without delays. In the last case, Requester 4 submits a data request which is approved with the required audit logs in place, and hence shared data main- tains its value. In the authorized access cases, protected data sharing can be made more efficient if health information sharing processes are designed with techniques that promote cooperation and distributed trust between multiple domain- s/entities. Effective techniques can ultimately enable auto- assurance [10] and auto-auditability [11], which reduces man- ual governance hurdles in health information sharing and in- creases the speed/volume of accessibility of protected data to authorized data consumers [12].

Blockchain technology can be a promising solution for creating effective techniques that can help in minimizing the above Loss of Value and the Loss of Opportunity is- sues in health information systems. However, a solution with Blockchain technology should address the following research questions: How to convey diverse datasets and overcome the lack of trust between parties involved in health data sharing? How can the health records be tamper-proof and distributed to allow for a faster data accessibility? How to enable a con- sistent representation of authorization to access health infor- mation in a secured network platform? How to create a smart platform equipped with a conversational agent i.e., a *chat- bot* that interacts with the data consumer with a knowledge- driven capability to have a better consumer-platform inter- action? How can the quality of health information service be improved by incentives for data consumers and providers to cooperate?

In this paper, we propose a novel "HonestChain" system that uses Blockchain technology to allow organizations to have incentive-based and trustworthy cooperation to either access or provide protected healthcare records. More specifically, HonestChain addresses the above research questions and expedites the protected data request validation process while ensuring auditability via a consortium Blockchain reference architecture and a chatbot-based user interface. We adopt a permissioned/consortium Blockchain architecture ow- ing to the following benefits that include: (a) relatively short deployment duration and less resource-intensive properties in terms of the consensus mechanism, and (b) more effective than a permissionless/public Blockchain architecture for protected data sharing amongst a consortium of organizations. Moreover, in the permissionless Blockchain architecture (using e.g., Ethereum), there is no barrier of participating entities as the access roles are open for anyone to join and a rigid consensus mechanism is ensured.

Lastly, we implement our HonestChain system using the Hyperledger Composer in a realistic simulation testbed [13]. Our experimental testbed is realistic because it comprises of a consortium of multiple domains, including a number of service provider peers cooperating in order to share pro- tected health data, as well as a set of requesters needing access to protected health records. Through testbed exper- iments, we compare the service time benefits with minimal request resubmission rate based on the reputation scheme in our HonestChain with state-of-the-art requester reputa- tion schemes such as: Recency-based [14], Catalog [15] and Manual [16]. The Recency-based scheme includes the in- formation about the last two requests submitted, whereas the Catalog includes requests that prompt users to choose particular dataset(s) from a limited catalog list; the Manual scheme involves filling out paper forms for data request and authorization/access processes.

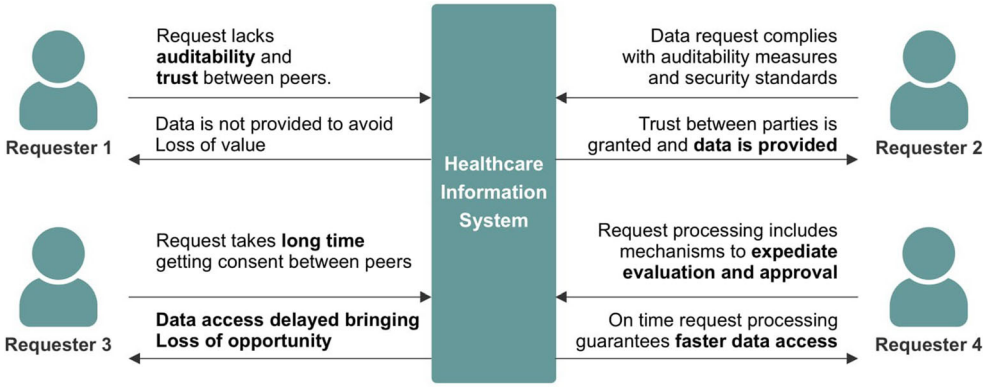

Fig. 1 Healthcare information sharing cases involving either authorized access or access with Loss of Opportunity/Value

The main contributions summary of this paper is as follows:

- We propose a consortium Blockchain based "HonestChain" system for real-time health information sharing that ensures auto-auditability and auto-assurance which reduces the Loss of Value and Loss of Opportunity issues.
- We equip our HonestChain system to facilitate distributed trust through analysis of User Reputation and Provider Reputation of peers to help a data consumer to request protected health datasets(s), and data provider(s) to provide health records in a secured manner.
- We include Chatbot assistance to guide users in submitting data request forms intelligently by minimizing errors, expediting data approval process and increasing reputation of the users.
- We evaluate our HonestChain implementation through comparisons with state-of-the-art schemes for data request process determination. Our results show that Honest Chain outperforms the existing schemes by providing least request resubmission rate through enforcement of chain code-based risk policies. Additionally, we demonstrate Block chain scalability capability through analysis of transaction rates for servicing a number of requests.

The remainder of the paper is organized as follows: Section 2 presents a background and relevance of Blockchain to the work in this paper. Section 3 details related works. Section 4 presents our HonestChain system design and components within a consortium Blockchain reference architecture. Section 5 describes performance evaluation results. Sec- tion 6 concludes the paper.

2 Blockchain background and relevance

Blockchain was first proposed in 2008 and implemented in 2009 [17]. A Blockchain is a distributed ledger that is managed by a peer-to-peer network. Blockchain can be regarded as a public ledger in which all committed transactions are stored in the ledger. The chain grows as new blocks or trans- actions are added. Transactions are validated and recorded by distributed consensus in the peer-to-peer network, elimi- nating a trusted central entity's need. Once transactions are validated, they become irreversible, verifiable, permanent, and secure on the Blockchain. Blockchain offers key char- acteristics, such as audibility, persistency, decentralization. Due to its advantages, it is widely utilized in applications such as finance, healthcare, transportation, and law.

It facilitates trust through its distributed consensus mechanism. Trust building in the HonestChain is performed through a peer reputation system that acts as an overlay on top of the Blockchain distributed trust capabilities. The reputation system helps us to determine the effectiveness of a service delivery performed by data custodian peers when a service request is submitted to the consortium by peer data requesters. Our consortium Blockchain trust building between data cus- todians and data requesters is further supported via chat- bot guidance that helps requesters to: (a) comply with data access standards by minimizing human errors, and (b) allows them to gain reputation in a consortium for expedited data access in future transactions. HonestChain's reputation scheme allows for creation and sustenance of the consortium with peers, and uses novel protocols to rate the peers objectively/subjectively using 'Requester Reputation' and 'Provider Reputation' metrics.

Our Honest Chain system leverages Blockchain platform and chat bot technologies to enable health information sharing in a secured, expedited and standards-compliant manner. Using a consortium Block chain-based approach, pro- tected data sharing is efficiently facilitated in HonestChain by using reputation value calculations of the peers (both Requesters and Providers) and by performing risk assessment of each transaction using automation to ensure *auto- assurance* and *auto-auditability*. The extent of benefit of Honest Chain system is dependent on the automation of dis- tributed trust. Additionally, usability of the chatbot-based re- quester guidance can also either be beneficial or can be a hinderence in the process of minimizing the Loss of Value and Loss of Opportunity issues. Further, our Honest Chain system provides access to protected data sets, but their analysis will require integration of various analytic tools and visualizations by users such as e.g., Jupyter notebooks.

3 Related works

3.1 Health information sharing systems

Health data sharing challenges have been widely approached by organizations for many years around efficient and secure ways to conduct the data brokering process. The data brokering process includes: data integration, data protection, Institutional Review Board (IRB) approval, brokering auditability, data assurance, and data request for analysis/visualization.

The data integration issue has been an extensively studied topic in prior works. Related efforts focus on defining frameworks to consolidate healthcare data from disparate sources into a unified platform. Authors in [18] proposed a common data model to serve as data hub for data brokering processes, which improves data accessibility and availability. Works such as [19] that aim to improve data protection automate the data de-identification process and

centralize the IRB request evaluation. While above works expedite the data brokering function, all of them however require human intervention to evaluate requests and manually approve or deny them. Moreover, they don't systematically address the issues related to auditability and assurance of the brokering process, and don't devise methods to minimize human intervention. In contrast, there have been recent efforts such as the work in [20] to semi-automate the honest broker processes through automation in compliance checking to expedite the data sharing.

Our goal in the current work is to extend the above prior works and implement a fully automated honest broker solution using Block chain and chat bot technologies with minimal custodian-in-the-loop intervention. Our approach establishes distributed trust by improving efficiency in compliance checking, incorporating auditability, and including a common data model to improve data accessibility. Thus, we address long delays in data accessibility due to human intervention based on automation of assurance and auditability steps during requester-custodian co-operation in protected data access.

3.2 Guided data brokering with trust

The work in [20] proposed a semi-automated honest broker that partially addresses the lack of trust between data custodians and data consumers in order to improve the data sharing process. Their methodology does not focus on improving the consumer interactions within a trusted health information sharing platform. Trust can be established by having data custodians use guided interfaces such as conversational agents i.e., chat bots to avoid human errors in over/under- provisioning of data requests or enable quick submission of protected complex data requests. Advantages of using chat- bots have been presented in works such as [21], where mobile health care services have been improved using relevant knowledge bases to provide fast requirement analysis and quick response to address conditions of patients impacted by accidents.

However, the design of chat bot guided systems need to be built in a manner that ensures maximum service, component re-usability and scalability. In addition, chat bots need to be designed with suitable natural language processing (NLP) techniques as detailed in [22] based on ranking and sentence similarity calculations. Some of the algorithms in [23, 24] completed certain image processing operations with improved performance as part of providing users a better experience.

The novelty of our approach is in the development of an automated brokering system using a chatbot incorporated with the necessary NLP techniques that helps in improving the data custodians' and data consumers' reputation. Our chat bot development follows the best practices for chatbot creation as outlined in [25]. We leverage chat bot technology to minimize the service times caused by human errors and its integration with the Block chain technology minimizes trust bottlenecks.

3.3 Blockchain in broker systems

Several prior works address the problem of lack of trust in sharing protected healthcare data. One of the exemplar works in [26] proposes a trust-building brokering architecture that fosters patient-centric cloud eHealth services. This model seeks user feedback and enables auditability by tracking transactions through a Blockchain solution. Addition- ally, brokering systems with Blockchain technology can both improve the quality of patient care and reduce the cost of care with targeted safe sharing of healthcare data as shown in [27]. To overcome the limitations in a centralized architecture of health information sharing such as high dependence on network connectivity and a single point of failure, authors in [28] propose the use of a Blockchain solution. Their approach uses distributed ledger technologies to facilitate multi-site, collaborative studies in the data-intensive sciences such as genetics/genomics, and enables auditability through single institutional ethics review in their Blockchain platform. The work in [29] uses technologies such artificial intelligence, machine learning and Blockchain to enable re- searchers to access medical data by transforming simple facial pictures and videos into powerful sources of data via predictive analytics.

Blockchain technology when integrated with an online machine learning model as shown in [30] can further help in distribution of partial models, and also drive the design new proof-of-information algorithms. Blockchain combined with asynchronous collaborative machine learning can help with sharing of information between distributed agents [31]. Blockchain can be utilized for efficient and privacy-preserving data sharing. In this context, work in [32] allows distributed management of identity and authorization policies by lever- aging Blockchain technology. Similarly, in [33], Blockchain is used for the identity authentication, and the access control permission of data is redesigned and stored on the Blockchain. Authors in [34] use Blockchain to extract and analyze the re- quested permissions in an Android application. Work in [35] proposed a Blockchain-based and decentralized trusted service mechanism for the crowdsourcing system in 5G-enabled smart cities to overcome security risks in crowdsourcing systems. Furthermore, a Blockchain-enhanced security access control scheme is shown to help support traceability and revocability within IoT systems in [36]. Blockchain technology has also been leveraged to ensure data integrity [37] in cloud applications as part of cloud audit systems. To reduce storage and communication cost, works in [38, 39] propose ultra-lightweight mutual authentication protocols. A salient work in [40] aims to improve the process quality by providing published results against COVID-19 on the Blockchain, and thus improves productivity of scientific research for the researchers.

Our HonestChain is inspired by the above works and is a system that takes into account both the objective and sub-

jective reputation attributes. Our reputation scheme with an automated risk assessment technique ensures that data requests comply with health information sharing standards. Through our Blockchain based platform, we incentivize the consortium of peers through rewards that use the reputa- tion of data custodians and data consumers. This interaction between data custodians and data consumers in our HonestChain builds a trusted network of peers e.g., data custodians automate data requests via audit log notifications to data custodians, and serve data consumers with a faster data access decision process.

4 HonestChain system design

4.1 HonestChain system overview

Figure 2 illustrates our proposed reference architecture in a consortium where our HonestChain is hosted on a cloud infrastructure that is accessible by different peers that want to leverage the service. The key component in our HonestChain is the consortium Blockchain-based trust setup built on top of our reputation scheme, and incentives for information sharing as detailed later in this section. Within this consortium, we assume that there are peers (Requesters) re- questing for protected health data, and peers (Providers) pro- viding the records from cooperating domains. Furthermore, peers in each domain are assigned with a reputation value based on their contributions to the other consortium peers.

Our HonestChain rates the Requester and Provider peers using metrics such as compliance score, dataset risk, and user's feedback. Using our reputation scheme, we minimize the is- sues of Loss of Value and Loss of Opportunity in enabling protected data access. HonestChain system design includes on-chain and off-chain components for storage, processing and sharing of health information:

On-Chain: this component fetches and displays the details such as e.g., `user id`, `dataset id`, `risk level`, `decision`, `reputation` from the Common Data Model (CDM) and the related automated honest broker services in HonestChain. These details are fed into the chaincode that helps in calculation of User and Provider Reputations. Hyperledger Fabric platform often uses the terms smart contract and chain code interchangeably. When chain codes are deployed, all smart contracts within it are made available to applications.

Off-Chain: this component stores information such as the details about the domain form filled by requester, the re- quester details, compliance score, and dataset details. De- pending on the number of requests submitted and the het- erogeneity of requested data, the storage of the related data will require large amounts of storage (in the order of tera bytes or even peta bytes in core network domain scenarios) and a homogenized data format. For this purpose, we utilize the CDM as an off-chain storage that interacts periodically through the related honest broker services in HonestChain. The `dataset id` from the CDM is fetched from honest broker services and is referenced in our chaincode. This approach allows us to deliver a dynamic, standards- compliant and efficient protected data retrieval process in a peer-to-peer manner.

4.2 HonestChain system architecture

Our HonestChain architecture includes the CDM module, and a cloud instance hosting the honest broker services, the User Interface (UI), the Blockchain module, and computa- tional analytics workspace module as shown in Fig. 3. The CDM module integrates heterogeneous data from mul- tiple sources

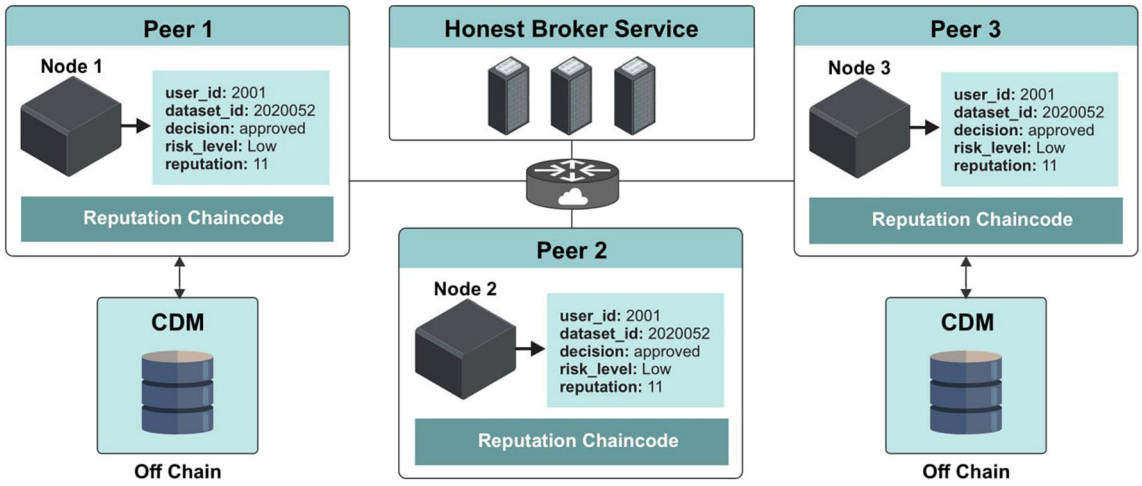

Fig. 2 Honestchain reference architecture that features on-chain/off-chain components within a consortium of peers involving an honest broker service, dedicated blockchain nodes with Hyperledger configurations, chaincodes and CDM

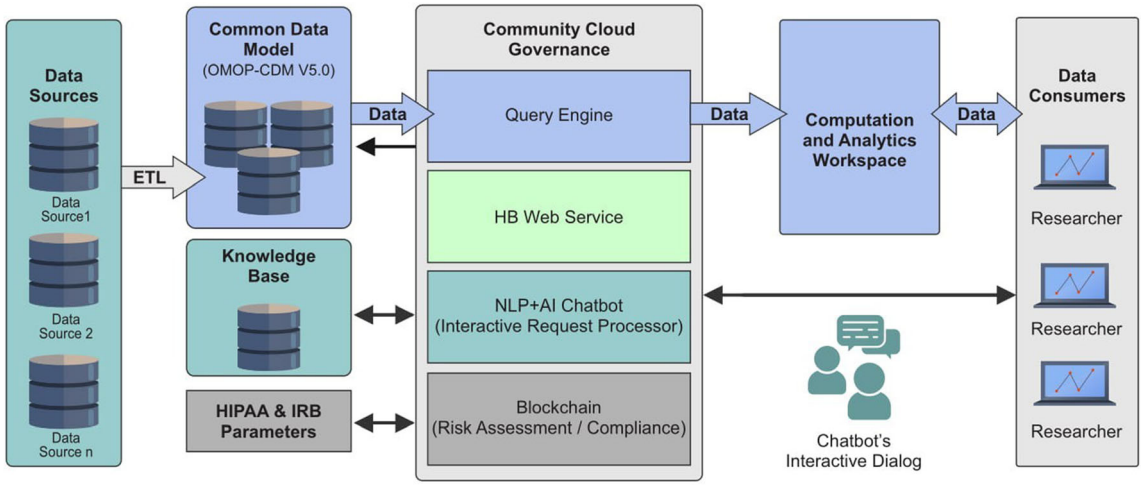

Fig. 3 Functional modules of HonestChain that include: Blockchain module for risk assessment and transaction tracking; the guided chatbot UI for data request processing; the Honest Broker service that mediates the Blockchain, and UI

into a common model based on the OMOP- CDM framework [2]. An Extraction Transformation and Load (ETL) process retrieves the data from each data-source and applies the required data transformations to homogenize and fit the data into the common model. Once the data is in the CDM module, standard queries can be defined and executed to quickly retrieve the data in a consistent manner. The CDM module alleviates the users from the need to know specific data query languages, or database structures.

To allow the Requester to interact with the HonestChain system, the UI module provides the functionality to submit and review the status of data requests. System administrators can use the UI to review the status and manage data requests submitted by Requesters. The UI provides data re- quest forms to define the data needs in terms of data Identifiers and data Domains. It also provides a dashboard for the Requesters to review the status of requests and retrieve their dataset, if the request has been approved. If the Re- quester needs assistance to fill the data request form, a chat- bot service in HonestChain provides the necessary guidance for the Requester to answer the questions and make the right item selections in the request and domain forms. The chatbot service relies on the information from a knowledge base to generate the conversation responses. Once a request is sub- mitted via the UI, the Blockchain module, based on the Hyperledger Composer, calculates the risk and verifies if the request is in compliance with the HIPAA and IRB parameters previously approved and stored in a relational repository. We remark that our Blockchain implementation helps to enforce HIPAA privacy rules through risk calculations for data requests based on factors such as e.g., healthcare data sensitivity levels and data requester reputation. In this way, HonestChain ensures that individually identifiable information release can be automated when the calculated risk is low, and routed to a governance committee for other risk cases.

Based on the calculated risk level and compliance checking, the request is approved or denied and the result is recorded in a Blockchain node. Once a data request is approved, the request record includes a link that allows the Requester to trigger a query, and retrieve the data in the CDM format. Once the data is retrieved, it is mabe available in a Computational and Analytics Workspace module for the Requesters to consume the data in their analyses/ visualizations.

4.3 HonestChain system modules

Herein, we present an example to show how to leverage our HonestChain system on a request for health data from data custodians. We can identify two main steps in the data brokering process as shown in Fig. 4. The step (A) corresponds to request submissions following the IRB and HIPAA guidelines. Although it is expected that this step is completed in a single request, our solution provides the flexibility for the users to customize the request if needed. Consequently, the process will turn into a negotiation between the data custodians and the data requesters about the scope and conditions to release the data for a particular research/clinical project. The negotiation is an interactive process with one or more iterations until an agreement is met. Upon IRB and HIPAA compliance approval through a governance process, multiple data requests can be filed during the life cycle of a related research project. The request submitted is handled by step (B), in which a validation process is conducted to determine if the requested protected data is within the limits agreed during step (A). A validation and data request refinement could also take place in an interactive manner. Upon approval, the data is retrieved from the repository and the dataset is made available to be used by the data requester. HonestChain functionality aims to address step (B), and

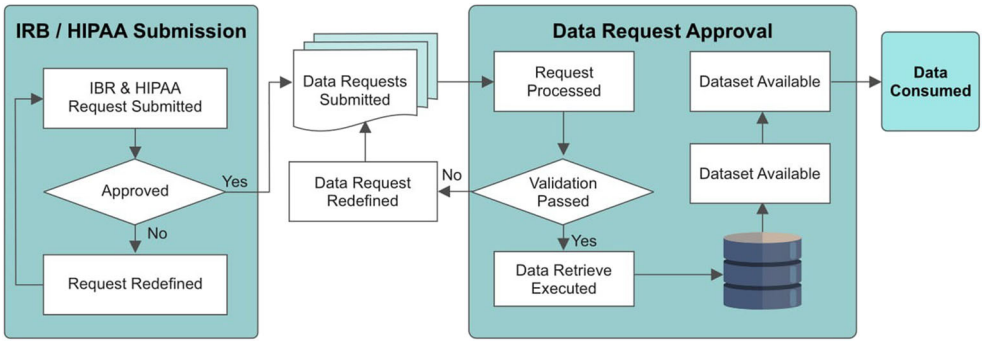

Fig. 4 Data Request Process with: (**a**) submission of the IRB and HIPAA compliant request, and (**b**) data request approval

assumes the IRB and HIPAA negotiation process has already been performed. As these two steps are automatically assisted by our HonestChain system, no significant delays are introduced while helping the user to complete the request to avoid Loss of Opportunity.

The sequence flow for system module interactions in HonestChain is illustrated in Fig. 5. Data requesters lo- gin via the user interface that handles the authentication and authorization step. The login credentials are then mapped to the tokens in the Hyperledger Composer. Upon successful authentication and authorization, the data requester fills the 'data request form' and the corresponding parameters are stored in a relational database. This is followed by submitting the transaction to the consortium Blockchain, which executes the relevant chaincode and broadcasts the decision to the consortium peers. If the data request is approved, a data query is supported followed by the related data retrieval for analysis/visualization. An alternate approach can involve storage of the custodian data in the InterPlanetary File System (IPFS) and maintenance of the references to the hashes in the chaincode. IPFS features as a high-performance and secure protocol for a peer-to-peer network to store and share data across a distributed file system.

4.4 HonestChain user interface

The HonestChain system provides an UI that allows users to submit requests, and a backend Blockchain process calculates compliance scores, and performs risk assessment. Based on these calculations, a reputation value is attributed to the data

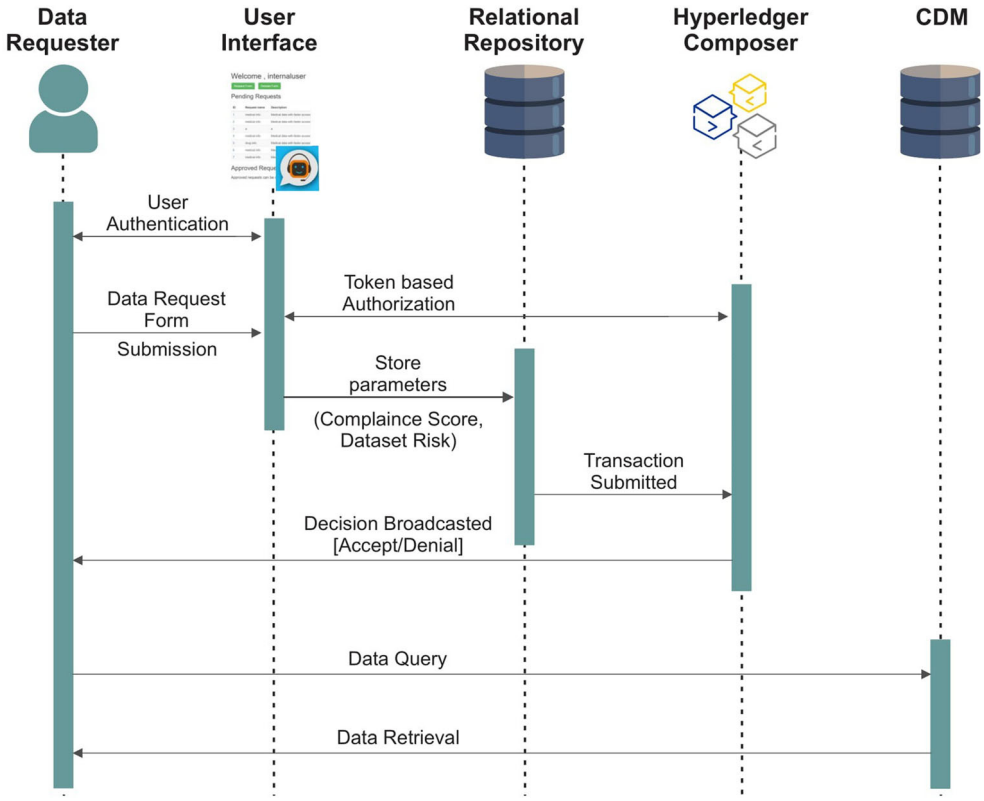

Fig. 5 Sequence diagram showing interaction flows between the HonestChain system modules

requester(s) and data provider(s), using which decisions for approval or denial are made for the protected data access.

Our UI is built using the Flask framework in the back- end, and the front-end is built using HTML, CSS, Boot strap and JavaScript. Additionally, our Blockchain module leverages the Hyperledger Composer deployed on a multi- organization setup on Hyperledger Fabric. Being a consortium Blockchain, it allows us to easily create business net- works that can be deployed on a public cloud infrastructure. Figure 6 provides an example scenario of our workflow instantiation. HonestChain allows internal (Intra-domain) and external (Inter-domain) Requesters to log in and complete a domain form and request form.

When a Requester login is authenticated in the dash- board, he/she is directed to complete a request form which consists of data request fields (e.g., project title, data set name), and the domain form which consists of project related information, domain fields (e.g., laboratory reports, radiology images, demographic information). The domain form allows users to specify the domain fields corresponding to the request data. We perform compliance checking where domain policies are compared with the domain form items submitted by the Requester. Once compliance verification is completed, the details are sent to the Blockchain module.

To fill the request and domain forms, the Requester is guided by our HonestChain chabot. Our chatbot is implemented using the Dialog flow API maintained by Google to promote human–computer interaction based on natural language conversations. Our HonestChain user interface hosts the client API for Dialog flow and is built using the Flask framework on the back-end and HTML, CSS, Bootstrap and JavaScript in the front-end. Whenever a requester asks his/her question, the request-data intent is sent to the chatbot through API calls and this will fetch a response from the knowledge base which supports collection, retrieval and sharing of prior curated knowledge. Thereby, the chatbot guides the Requesters with pertinent answers using the knowledge base, and assists them e.g., when they are facing difficulty in filling out the request or domain forms.

The Blockchain module is triggered to ensure auditability, transparency and immutability. The interaction of the UI with the Blockchain network starts with the transfer of data request parameters from the REST API calls as shown in Fig. 7. We map the requesters' identities through the UI with our Blockchain network peer interactions. Through authorization and authentication by the Hyperledger Fabric, certificates are generated and the peer is authorized to instantiate the chaincode. Our Blockchain network is comprised of the Hyperledger Composer model file, script file, access control language (ACL) file that provides declarative access control over the elements of the domain model, and query file. All of these files make the Blockchain Network implementation configurable in the Hyperledger Fabric.

The chaincode in Hyperledger Composer allows approval or denial of a request. The chaincode implementation is essentially a script file that is added on to the distributed nodes.

To perform decision making, several parameters are retrieved through the REST API calls i.e., user id(I_d), compliance score(C_s), and dataset risk. This step is followed by consensus and deployment of the chaincode, which ultimately adds the transaction to the global transaction list. The consensus here i.e., Proof of Authorization occurs very fast and the transaction is successfully submitted. Upon submission, data custodians can inspect the Blockchain to view all the transactions and view the risk levels, reputation values and data access decisions for auditability. This information is broadcasted to all the relevant peers (other data custodians) in the network. Our HonestChain system also provides the functionality to automatically redirect the response to the UI, where a Requester is able to see the decision as approved or denied.

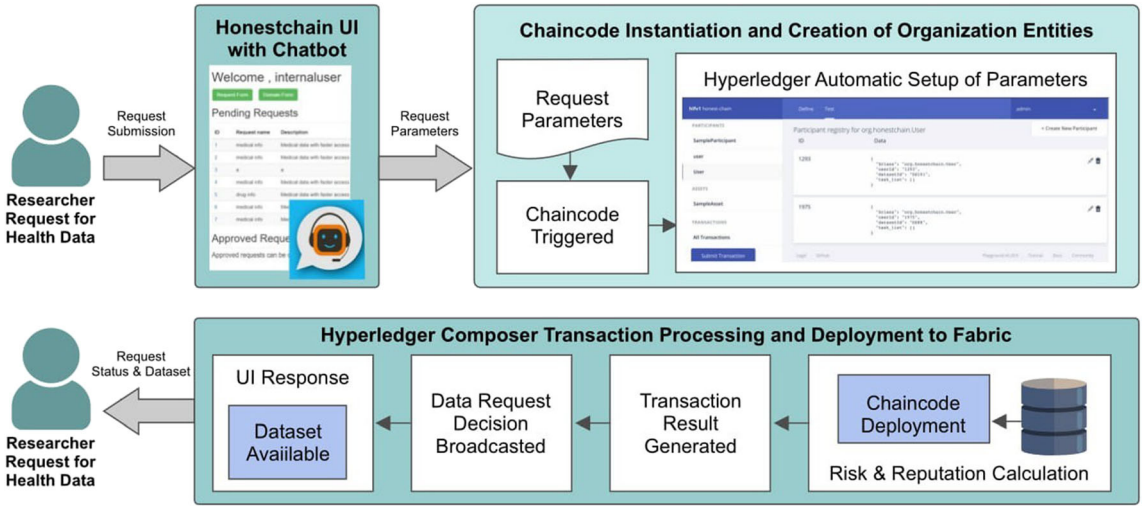

Fig. 6 HonestChain workflow with requester peers and provider peers sharing health information cooperatively

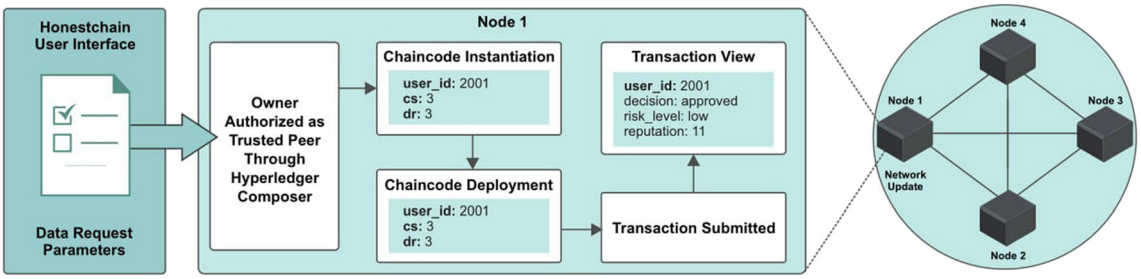

Fig. 7 HonestChain system showing chaincode deployment and transaction submission stages over a distributed network

With the approval decision, a call to the CDM module takes place that fetches the data and is transferred back to the UI. All of these REST API calls occur instantly and the requester is able to view the requested health data in a few seconds.

4.5 HonestChain system roles

There are three different roles i.e., Healthcare Data Requester, Healthcare Data Provider, Healthcare Brokering System Administrator that serve as central actors in our HonestChain implementation. In the following, we provide their definitions:

4.5.1 Healthcare data requester

Requester is an actor who needs the health data and submits the requests to a consortium of peers. Once the request is identified, our honest broker service in HonestChain determines the parameters to send to the Blockchain network, and determines the Requester's reputation. The transaction gets submitted when all the parameters are sent via REST API calls and the Provider is notified.

4.5.2 Healthcare data provider

Provider is an actor who provides health records to the re- quests from a Requester if the decision is approved from the Blockchain chaincode. To incentivize Providers to provide high-quality services, a reputation R_p is given to the Provider so that their service is regarded as a trustworthy and reputable service.

4.5.3 Healthcare brokering system administrator

Admin is an actor who: (a) analyzes and monitors the sub- mission of the data request, and (b) identifies the result of the request to determine if any manual action of reviving and ap- proving the request is needed. The Admin is responsible for taking care of the 'manual approval required' decisions by passing the details of the requester and requested data to the honest broker governance process (manual or automated) for further evaluation. Admin here is also the point of contact for further assistance for the request related queries by the Requesters.

4.6 Reputation based healthcare data brokering protocols

4.6.1 User reputation

Prior works in [41] [42] [43] determine HIPAA compliance in order to consider functionality of healthcare systems based on predefined regulations. Our proposed HonestChain solution features an automated HIPAA [1] compliance checking method to establish the trust and ensures auditability through Blockchain of data request transactions. To deter- mine the trust of a Requester, we first calculate if the data requested by the Requester is compliant or not. This process is done by comparing the policies of requested data with the answers from the data domain request form filled by Requester. A Requester needs to fill the domain form that consists of 18 questions. Each domain fields include *Yes* (*Y*), and *No* (*N*). Additionally, each type of data (aggregated, de-identified, limited, or identified) also maintains its own policies based on the sensitivity. The data policies are also stored in the form of a list of 18 respective *categories*. Furthermore, we categorize the policies based on the Requester role that is internal and external. The data policies include 1 and 0, where 1 means compliant and 0 means not-compliant.

HonestChain retrieves the answers from the form and compares them to the policies of the requested data. The recorded comparison outcomes are assigned a compliance score as shown in Eq. 1. The equation first computes the score and then the resulting value is normalized in the range from [1, 10] as done by Eq. 2. We compute the risk on the basis of output of the average function as shown in Table 2. This function takes as an input compliance and dataset risk. Once the average is computed, the risk levels are computed and are described as Low (L), Medium (M) and High (H) risks as shown in Eq. 3. Our reputation scheme is based on two modules that are reputation of Requester and reputation of Provider. The reputation of the Requester is computed based on the risk levels. We describe our reputation as 0, −1 and + 1 based on the risk levels as shown in Eq. 4. The base reputation default is set to 10 for a new Requester, and this value is updated based on new requests submitted.

Table 1 shows the notations used in the remainder of this section. The Score calculation (S_i), Approved Data Element (a_i), Requested Data Element (r_i), Compliance score (C_s), Risk evaluation (R_u), and Reputation value calculation (B_w) can be given as (Table 2):

Score calculation: where ($0 < i < n$), and n = number

$$S_i = \begin{cases} 1 \text{ if } (a_i = r_i) \text{ or } (a_i = 1 \text{ and } r_i = 0) \\ 0 \text{ if } (a_i = 0) \text{ and } (r_i = 1) \end{cases} \quad (1)$$

of terms in domain form, $(a_i) = 0$ if it was not included in the request and $(a_i) = 1$ if it was included, and $(r_i) = 0$ if it was not included in the request, and $(r_i) = 1$ if it was included.

Compliance score normalization from 1 to 10:

$$C_s = \left[\left(\sum_{i=1}^{n} (S_i * 10) \right) / n \right] \quad (2)$$

User Risk Level definition:

$$R_u \begin{cases} L : C_s \in [1, 3] \\ M : C_s \in [4, 6] \\ H : C_s \in [7, 10] \end{cases} \quad (3)$$

User Reputation calculation:

Algorithm 1 shows the compliance score computation steps

$$B_u \left\{ \begin{array}{c} -1 \text{ if } (R_u = H); \textit{request is denied} \\ 0 \text{ if } (R_u = M); \textit{manually evaluated} \\ 1 \text{ if } (R_u = L); \textit{request is approved} \end{array} \right\} \quad (4)$$

Table 1 Notations used in the descriptions

Notation	Description
C_s	Compliance score
R_u	Risk associated with requester user
S_i	Score calculation
L	Low risk
M	Medium risk
H	High risk
n	Number of interactions
β_k	Average reputation
β_w	Overall reputation
B_u	Requester User Reputation
B_p	Provider Reputation
S_p	Provider score given on each feedback
K	Initial provider reputation
I_d	User Id

for a number of requests submitted to the HonestChain system. The main purpose of the steps is to utilize the User Reputation protocol for computation of risk levels. As an input, the chaincode receives various requests submitted by users. For each r_i, the compliance score is calculated which is then used to obtain the risk level. Following this, we use s_i and n to determine c_s.

Algorithm 1 Algorithm for calculation of the Compliance Score that is used to obtain the Risk Level

Input: r_i List of Requests
Output: c_s Compliance Score
Function `Main` (r_i) **:**
  foreach r_i **do**
    if $r_i = a_i \vee a_i = 1 \vee r_i = 0$ **then**
      $s_i \longleftarrow 1$;
    if $a_i = 0 \wedge r_i = 1$ **then**
      $s_i \longleftarrow 0$;
  $c_s \longleftarrow$compliance (s_i, n);
End Function
Function *Compliance* (s_i, n) **:**
  $c_s \longleftarrow 0$;
  foreach s_i **do**
    $c_s \longleftarrow (c_s + s_i * 10)/n$;
  return c_s;
End Function

4.6.2 Data provider reputation

Chatbot guidance in the request form helps the Requesters to fill the data more accurately, which in return increases the reputation of the Requester. After the request is handled successfully and protected data access is granted, the Requester will fill out a feedback form where they subjectively give responses on Provider's performance. In this way, we account for subjective opinions in increasing Providers' reputation in addition to using objective metrics such as number of requests handled, service time per request, number of feed- back received, etc. We calculate reputation of the Provider based on Eq. 5 borrowed from a related work [44]: In our consortium Blockchain, we ensure that the providers

$$B_p = \left\{ \begin{array}{c} 1 \text{ if } (\textit{positive}); \textit{trust worthy} \\ 0 \text{ if } (\textit{neutral}); \textit{no assessment} \\ -1 \text{ if } (\textit{negative}); \textit{not trust worthy} \end{array} \right\} \quad (5)$$

are incentivized to provide the accurate data and the information is shared across the trusted peers. We incorporate an optimistic approach where we provide rewards to Providers as an incentive to share the protected healthcare data. This

Table 2 Risk Assessment for exemplar requests handled in the HonestChain system

Data source (A)	Data source (B)	Data source (C)	Overall likelihood (f1:max(A,B,C))	Compliance score (*Cs*)	Total risk score (f2=avg(f1, *Cs*)	User risk level (*Ru*)
3	2	3	3	3	3	L
5	6	4	6	6	6	M
8	9	10	10	10	8	H
2	5	9	9	6	8	H
2	4	6	6	6	6	M
3	5	6	6	3	5	M
3	8	9	9	3	6	M
5	6	10	10	10	10	H

reward is the increase in reputation that helps them by allowing more Requesters to be paired to use the Provider's data. Our reputation value calculation is given by:

$$B_p = K + \left(\sum_{i=1}^{m} S_p \right) \tag{6}$$

where: $(0 < i <= m)$; K is a constant that represents the initial reputation of provider p; m is the number of feedback values received on Provider p; S_p is the value given to Provider p on each feedback.

Initially, we assign a base reputation of 10 by default that is given by constant K to the Provider. Our base reputation is given in order for defining a specific matrix throughout out Through our reputation value calculation, providers can increase or decrease their reputation. The determination of reputation value is given by S_p, which is obtained through a feedback form provided to the Requester after the user receives the requested data. Requester can rate the Provider and the value goes into the −1, 0 or 1 category. The over- all feedback given by the Requester results in a cumulative value calculation for the Provider. In this optimistic model, our goal is to allow Provider and Requesters to gain reputa- tion by contributing honestly and sharing the healthcare data in a trusted, automated manner via an immutable ledger.

4.7 Incentives for sharing

We consider risk and reputation as the primary HonestChain factors in performing the decision making to authorize protected data access. This allows the HonestChain to rate a Requester and a Provider based on their respective historic feedback and risks. We follow a semi-legal approach including both objective and subjective ratings as detailed above, where we focus on determining the reputation of a Requester and Provider(s) based on their service performance and data request parameters. With the historic reputation information, and owing to the design of the reputation protocols in our HonestChain, we enable a trustworthy platform in HonestChain for health information sharing. A higher reputation value leads to a higher probability of a Provider peer being selected for delivering service to the Requester, and a higher reputation value leads to a higher probability of Requester peer data being approved and delivered in a fast manner in the future.

4.8 Exception Handling.

Prevention against Sybil attacks Sybil attacks occur when the attacker disguises as an authorized user and generates multiple illegitimate and fake identities in order to disrupt the functioning of the service and to take undesired con- trol over the peers within the consortium. Our HonestChain system allows an inbuilt trust creation through certificates. When Hyperledger Composer is deployed to the Fabric, all the Hyperledger Fabric Certificate Authority servers share the same database for keeping track of identities and certifi- cates. The identity management here is centralized and helps in protection against Sybil attacks.

Replay attacks Replay attack occurs when a user tries to delay or repeat the data transmission in a network. Our scheme is immune to replay attacks due to the incorporation of Admin authorization. It allows us to associate a certificate and a private key. The ability to deploy a network is only given to the trusted authority and furthermore, each transaction has a unique timestamp id that can be traced back to the user's identity. Additionally, due to automatic generation of a valid private key, a user without a key is unable to modify the content and the transaction will not reach consensus.

Authorization In our scheme, we first authorize the Re- quester through our UI and then we map his/her identity to the Blockchain chaincode. Our scheme only allows the trusted authorities to execute transactions. If the connection profile matches with the peer's details, then the chaincode is

executed. This mapping allows us to trace-back the real identity of Requesters to avoid potential frauds and thefts. Additionally, mapping of the user credentials with the Hyperledger Composer peers prevents the creation of multiple fake identities.

5 Performance evaluation

5.1 Experiment Testbed setup

To evaluate our Honestchain, we implemented our solution using a realistic simulation testbed on a public cloud infrastructure as shown in Fig. 8. In this testbed setup, we included a node dedicated for the User Interface (UI). Our UI is built using the Flask framework and is integrated with a chatbot created using Dialog Flow following the best practices [25]. Requester fills the protected data access request in "request form" with the guidance of the chatbot. The Blockchain implementation is hosted on an AWS EC2 instance and involves an Hyperledger Composer installation that utilizes 20 GB memory. Additionally, we utilize an AWS RDS instance for hosting the CDM service. To allow CDM service to handle large datasets, we configure 40 GB memory on the RDS. The request details are send to the HonestChain services for risk assessment calculations, and the peer reputation calculations. Based on the calculated values, HonestChain allows data custodians to automate decisions on approvals or denials of the data requests.

We considered state-of-the-art healthcare requester reputation schemes to compare performance and evaluate reputation values of our HonestChain. Specifically, we compare the performance of HonestChain with Recency-Based [14], Catalog [15], and Manual [16] schemes. The Recency-based scheme includes the information about the last two requests submitted, whereas the Catalog includes requests that prompt users to choose particular dataset(s) from a limited catalog list; the Manual scheme involves filling out paper forms for data request and authorization/access processes. We choose these schemes because they help us in deriving close-to-real results based on the types of request submissions.

We performed simulation experiments by choosing a different set of Requesters and Providers over 25 iterations. The goal here was to simulate real world situations that al- low us to create a fair chance of interactions. Each Requester and Provider have different values of data corresponding to the dataset that they request, or the dataset that they pro- vide. We performed simulations of the request process by taking into account parameters such as: type of dataset required, compliance score, and dataset risk levels to deter- mine the reputation values. Each request associated with these schemes have different values of data corresponding to the data request parameters that they employ. We evaluate our HonestChain in the testbed experiments using metrics such as: reputation values, service time and request resub- mission rate. Reputation value is the cumulative score as- signed to the Providers/Requesters based on the approval/− denial of a series of requests. Service time is the total time that our HonestChain takes to process each request. Request resubmission rate is the number of requests that are resub- mitted upon denial in the previous transaction.

5.2 Reputation scheme results

Our first experiment was to evaluate the decision making process in pairing a Provider, when a request arrives from a Requester for a particular dataset. We performed simulations of Requesters and Providers for evaluating decision making schemes. We simulated a total set of 20 Requesters and Providers. From this total set, we randomly generated a subset of 15 requests evaluating the various decision making schemes. With our HonestChain, the Requesters ended up pairing with Provider(s) based on their calculated Reputation values. The other schemes use different algorithms for pairing the Requesters and Providers. For instance, the Recency-based scheme will pair one Requester or Provider based on the last couple of requests, whereas the Catalog scheme will pair the Requester/Provider peers based on the request submission involving the completion of a pre-defined form (without guidance); the Manual-based scheme will pair the Requester/Provider through a long tedious form submission process. Our results show that Requesters and Providers paired by our HonestChain have better overall reputation values when compared to the state-of-the-art schemes. This can be seen in Fig. 9(a), where the HonestChain's performance improvement

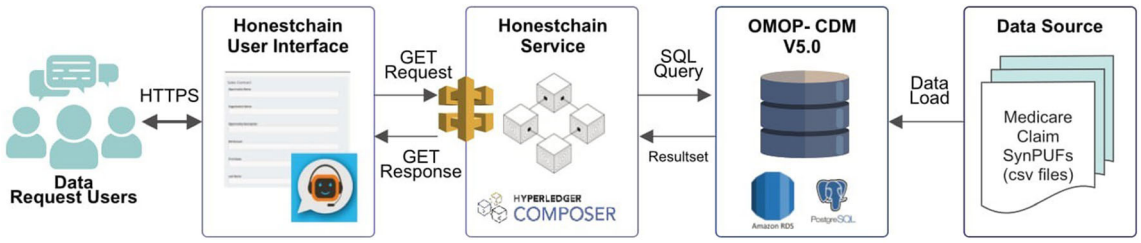

Fig. 8 Cloud testbed used to evaluate HonestChain system performance with experiments involving a distributed network

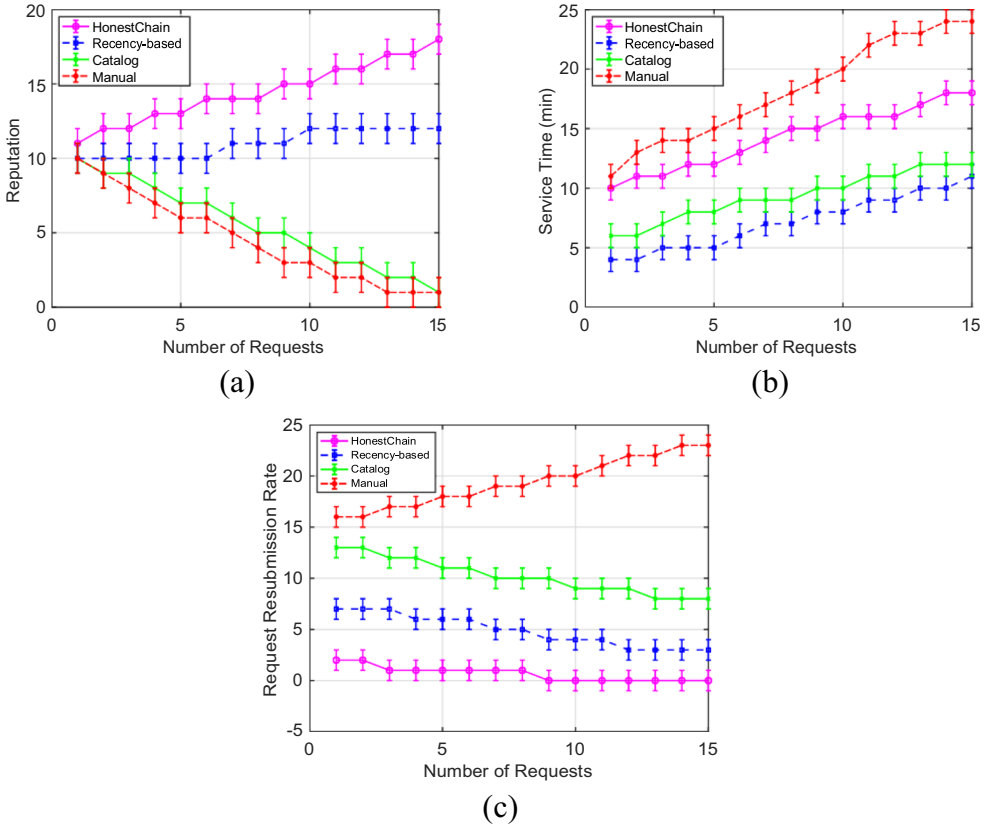

Fig. 9 HonestChain performance comparison with state-of-the-art schemes for: (**a**) reputation values for requesters, (**b**) service times, and (**c**) studying their performance trade-offs in terms of request resubmission rate

ranges from 1.3x - 4x times higher in terms of reputation values. This improvement in reputation in our HonestChain is due to the fact that we consider a comprehensive set of parameters to determine the Reputation values, rather than pairing Requesters/Providers based on a pre-defined form filling criteria in Catalog or using last couple of submitted requests in Recency-based approach. By comparing performances in the reputation of different state-of-the-art schemes, we were able to carry out step of analyzing performance trade-offs in service time and request resubmission rate.

Upon choosing the Provider using our HonestChain, we analyze the performance trade-offs in the service time, with the request resubmission metric. Our simulations helped us in determining performance and tradeoffs in different state- of-the art schemes. As shown in Fig. 9b and c, HonestChain takes up to 1.5 times more service time in the worst case as compared to the Recency-based [14], Catalog [15], and Manual [16] schemes. This is due to the rigorous mul- tiple stages involved in the HonestChain process of compliance checking, risk determination and decision making process and data retrieval process using the CDM module. However, these steps only consume a few minutes and the related overhead times can be compensated by using automation to make the risk assessment strategies more efficient.

More importantly, we should note that our HonestChain produces the least request resubmission rate compared to the other schemes. It is due to our chatbot guided interface, consortium Blockchain architecture and the automation for efficient risk enforcement policy implementation that together ensure that protected data requests are compliant with the security standards (thus minimizing Loss of Value) and have inherent auditability that avoids manual intervention (thus reducing the chance of Loss of Opportunity). The request resubmission rate is an important measurement of the healthcare broker systems, and all Requesters will inherently get a much higher weight or reputation in real-world scenarios if their resubmission rate is low. Thus, we show that our HonestChain has much higher performance overall considering trade-offs in comparison to the state-of-the-art decision-making schemes in pairing the appropriate Requesters and Providers in a consortium.

In Table 3, we summarize various properties about algorithms that we have used in the HonestChain system for comparison purposes. Based on HonestChain comparison with other schemes such as Recency-based, Catalog and Manual, we found notable strengths and weaknesses that can help in evaluating suitable models. HonestChain has an accurate assessment of reputation through its risk-based reputation scheme that outperforms the data brokering process in other schemes. Our scheme ensures that Requesters and Providers have better reputation due to our robust reputation based data brokering protocols. With the limited freedom for

Table 3 Properties comparison for different algorithms handled with the HonestChain system

Properties	Honest Chain	Recency based	Catalog	Manual
Features	Risk-based reputation scheme	Last two request reputation	Limited selection options	User-custodian compliance checking
Strength	Accurate assessment of reputation	probabilistic reputation	Suitable selection of data needs	Flexible process
Weaknesses	Limited freedom for ad-hoc requests	Limited view of overall performance	Lack of ad-hoc requests	Loss of Opportunity

ad-hoc requests to choose data sets present in the cata- log, we are able to ensure that the privacy of data requested is compliant with relevant health data standards.

5.3 Scalability results

Lastly, we evaluate our HonestChain based on the scalability performance as shown in Fig. 10. Specifically, we com- pare HonestChain throughput results involving measurement of transaction rate (throughput rate per second) in terms of increasing block sizes ranging from 10,000 to 100,000. For a fair comparison, we used the same transaction chaincode for all experiments. Each transaction process involves a series of independent read/write operations focusing on I/O, caching and parallelism. We get transaction throughput of 210,000 (tx/s) for a block size of 10,000 transactions. A smaller block size of 10,000 transactions corresponds to a lower batch size as per [45], which produces a lower throughput.

In contrast, a higher block size of 40,000 transactions corresponds to a higher batch size, which produces a larger throughput rate. As the block size further increases up to 40,000, we see the highest transaction throughput. When the block size goes beyond and reaches to a level of 60,000, there is a sharp decrease in the transaction throughput. This decrease can be attributed to the delay in accepting the committed transactions by the endorsing peers in the Hyperledger Fabric, and the limited number of endorsing peers to accept the consensus mechanism. We also note that this delay occurs due to the limited resources available in our cloud testbed setup to process transactions above 60,000. It is possible however to minimize this time delay by allowing participation of higher number endorser peers, and by increasing the resource availability in real-world HonestChain deployments to meet service demands. In summary, the scalability experiment results show that by achieving a highest throughput on 40,000 block size, our HonestChain system is practical and is able to scale for a reasonably large number of protected health data requests in a consortium of Re- questers and Providers.

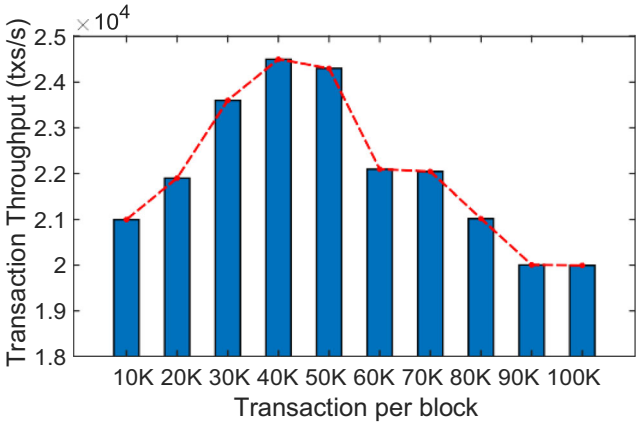

Fig. 10 HonestChain Throughput results involving measurement of transaction rate in terms of block sizes ranging from 10,000 to 100,000.

ions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not nec- essarily reflect the views of the National Science Founda- tion. The authors would like to thank the following colleagues who provided valuable inputs to this work: Matthew Joseph Carroll, Brett Young and Keaton Meyer

6 Conclusion

In this paper, we developed a novel HonestChain system that allows Providers to perform faster data decision making when processing protected data requests from Requesters. These Provider benefits for multi-source data sharing and analysis can support rapid innovations for clinical research informatics and engender next-generation decision support for researchers/clinicians in the cure of diseases.

Our evaluation results from a realistic experimental cloud testbed of a health information system show that our HonestChain is effective in increasing reputation of both Providers and Requesters. Consequently, HonestChain reduces the re- quest re-submission rate in comparison to state-of-the-art requester reputation schemes such as Recency-based, Cat- alog and Manual schemes that allow for secure and speedy access to protected data for authorized Requesters. Lastly, we showed that our HonestChain is practical and scalable to handle tens of thousands of transactions per block with high-performance.

As future work, HonestChain performance can be improved by integration of optimization strategies and related Blockchain policies for higher-scale workloads. Towards this aim, one can use high-performance computing back-ends in

cloud platforms as well as Jupyter notebook front-ends to enable faster data analysis/visualization for the requested protected datasets. Additionally, with a more robust testbed platform, we plan to conduct usability tests with real-world use case scenarios engaging key users in order to analyze transaction results in the HonestChain system. These results can subsequently be used to further improve the usability of the HonestChain system for diverse real data analysis use cases in healthcare data sharing applications.

Acknowledgments This work was supported in part by the National Science Foundation under award number OAC-1827177. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not nec- essarily reflect the views of the National Science Founda- tion. The authors would like to thank the following colleagues who provided valuable inputs to this work: Matthew Joseph Carroll, Brett Young and Keaton Meyer.

References

1. Hash J, Pauline B, Arnold J, Smith C, Steinberg D (2005) An introductory resource guide for implementing the health insurance portability and accountability act (HIPAA) security rule. US Department of Commerce, Technology Administration, National Institute of Standards and Technology
2. OMOP Common Data Model (CDM) V5.0. by Observational Health Data Sciences and Informatics (OHDSI) at: https://www.ohdsi.org. Accessed April 2021
3. Wu J, Guo S, Li J, Zeng D (2016) Big data meet green challenges: Big data toward green applications. IEEE Syst J 10(3):888–900
4. Wu J, Guo S, Li J, Zeng D (2016) Big data meet green challenges: Greening big data. IEEE Syst J 10(3):873–887
5. Atat R, Liu L, Wu J, Li G, Ye C, Yang Y (2018) Big data meet cyber-physical systems: a panoramic survey. IEEE Access 6: 73603–73636
6. Wu J, Guo S, Huang H, Liu W, Xiang Y (2018) Information and communications technologies for sustainable development goals: state-of-the-art, needs and perspectives. IEEE Commun Surv Tutor 20(3):2389–2406
7. Iroju O, Soriyan A, Gambo I, Olaleke J (2013) Interoperability in healthcare: benefits, challenges and resolutions. Int J Innov Appl Stud 3(1):262–270
8. Afshari H, Peng Q (2014) Challenges and solutions for lo- cation of healthcare facilities. Ind Eng Manag 3, 12(1)
9. Giannetsos T, Dimitriou T, Prasad NR (2011) People- centric sensing in assistive healthcare: privacy challenges and di- rections. Security and Communication Networks 4(11):1295–1307
10. Cannoy SD, Salam AF (2010) A framework for health care information assurance policy and compliance. Commun ACM 53(3): 126–131
11. King JT, Smith B, Williams L (2012, January). Modifying without a trace: general audit guidelines are inadequate for open-source electronic health record audit mechanisms. InProceedings of the 2nd ACM SIGHIT International Health Informatics Symposium 305–1314
12. Liang, X., Barua, M., et. al., 2012. Health share: achieving secure and privacy-preserving health information sharing through health social networks. Comput Commun, 35(15), pp.1910–1920
13. Berman M, Chase JS, Landweber L, Nakao A, Ott M, Raychaudhuri D, Ricci R, Seskar I (2014) GENI: a federated testbed for innovative network experiments. Comput Netw 61:5–23
14. Bourdev LD, inventor; Adobe Systems Inc, assignee (2008) Autocompleting form fields based on previously entered values. United States patent US 7,343,551
15. Kirby, J.C., Speltz, P., et. al., 2016. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. J Am Med Inform Assoc, 23(6), pp.1046–1052
16. Al Dogether, M., Al Muallem, et. al., 2016. The impact of automating laboratory request forms on the quality of healthcare services. J Infect Public Heal, 9(6), pp.749–756
17. Nakamoto S (2019) Bitcoin: a peer-to-peer electronic cash system. Manubot
18. Boyd AD, Hunscher et al (2005) The "Honest Broker" method of integrating interdisciplinary research data. In AMIA Annual Symposium Proceedings. Am Med Inf Assoc 2005:902
19. Felmeister AS, Masino et al (2016) The biorepository portal toolkit: an honest brokered, modular service oriented software tool set for biospecimen-driven translational research. BMC Genom 17(4):434
20. Valluripally S, Raju M et al (2019) Increasing protected data accessibility for age-related cataract research using a semi- automated honest broker. J Model Ophthalmol 2(3):115–132
21. Chung K, Park RC (2019) Chatbot-based heathcare service with a knowledge base for cloud computing. Clust Comput 22(1):1925–1937
22. Kavitha BR, Murthy CR (2019) Chatbot for health- care system using Artificial Intelligence. Int J Adv Res Ideas Innov Technol 5: 1304–1307
23. Li D, Deng L, Gupta BB, Wang H, Choi C (2019) A novel CNN based security guaranteed image watermarking generation scenario for smart city applications. Inf Sci 479:432–447
24. Zhang J, Yu K, Wen Z, Qi X, Paul AK (2021) 3D reconstruction for motion blurred images using deep learning-based intelligent systems. Comput Mater Continua 66(2):2087–2104
25. Lise Embley, technical writer, National Center for State Courts(2020). [online] Available at: https://www.ncsc.org//media/Files/PDF/About. Accessed April 2021
26. Kurdi H, Alsalamah S et al (2019) HealthyBroker: a trustworthy blockchain-based multi-cloud broker for patient-centered eHealth services. Electronics 8(6):602
27. Agbo CC, Mahmoud QH (2020) Blockchain in health- care. Int J Healthcare Inform Syst Inform 15(3):82–97
28. Rahimzadeh V (2018) Ethics governance outside the box: reimagining Blockchain as a policy tool to facilitate single ethics review and data sharing for the 'omics' sciences. Blockchain in Healthcare Today 1:1–0
29. Mamoshina P, Ojomoko L, Yanovich Y, Ostrovski et al (2018) Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical re- search and healthcare. Oncotarget 9(5):5665
30. Al Omar, A., Bhuiyan, M.Z.A., et. al., 2019. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Futur Gener Comput Syst, 95, pp.511–521
31. Hua G, Zhu L, Wu J, Shen C, Zhou L, Lin Q (2020) Blockchain-based federated learning for intelligent control in heavy haul railway. IEEE Access 8:176830–176839
32. Esposito C, Ficco M, Gupta BB (2021) Blockchain-based authentication and authorization for smart city applications. Inform Process Manag 58(2):102468
33. Shi N, Tan L, Li W, Qi X, Yu K (2020) A blockchain- empowered AAA scheme in the large-scale HetNet. Digital Communications and Networks
34. Ouaguid A, Abghour N, Ouzzif M (2018) A novel security framework for managing android permissions using blockchain technology. Int J Cloud Appl Comput (IJCAC) 8(1):55–79

35. Tan L, Xiao H, Yu K, Aloqaily M, Jararweh Y (2021) A Blockchain-empowered crowdsourcing system for 5G-enabled smart cities. Computer Standards Interfaces, p.103517
36. Yu KP, Tan L, Aloqaily M, Yang H, Jararweh Y (2021) Blockchain-enhanced data sharing with traceable and direct revocation in IIoT. IEEE Trans Industr Inform 1
37. Mohan AP, Gladston A (2020) Merkle tree and Blockchain- based cloud data auditing. Int J Cloud Appl Comput (IJCAC) 10(3):54–66
38. Al-Qerem A, Alauthman M, Almomani A, Gupta BB (2020) IoT transaction processing through cooperative concurrency control on fog–cloud computing environment. Soft Comput 24(8):5695–5711
39. Tewari A, Gupta BB (2017) Cryptanalysis of a novel ultra- lightweight mutual authentication protocol for IoT devices using RFID tags. J Supercomput 73(3):1085–1102
40. Yu K., Tan L, Shang X, Huang J, Srivastava G, Chatterjee P (2020) Efficient and privacy-preserving medical research support platform against COVID-19: a Blockchain-based Approach. IEEE Consum Electron Mag
41. Oh S, Cha J, Ji M, Kang et al (2015) Architecture design of healthcare software-as-a-service platform for cloud-based clinical decision support service. Healthcare Inform Res 21(2):102–110
42. IBM services: Getting your data ready for precision medicine. [Online] Available at: https://www.ibm.com
43. Community cloud architecure for Salesforce health care applications. [Online] Available at: https://www.salesforce.com/products/community-cloud/faq
44. Aberer K, Despotovic Z (2001, October) Managing trust in a peer-2-peer information system. In Proceedings of the tenth inter- national conference on information and knowledge management (pp. 310-317)
45. Gorenflo C, Lee S, Golab L, Keshav S (2019, May) Fast- fabric: scaling hyperledger fabric to 20,000 transactions per second. In 2019 IEEE international conference on Blockchain and Cryptocurrency (ICBC) (pp. 455-463). IEEE

Soumya Purohit received her BS degree in Information Technology from Banasthali University, Jaipur. She is currently pursuing her M.S. degree in the Department of Electrical Engineering and Computer Science at University of Missouri-Columbia. Her current research interests include Blockchain Technology and Cloud Computing.

Prasad Calyam received his MS and PhD degrees from the Department of Electrical and Computer Engineering at The Ohio State University in 2002 and 2007, respectively. He is currently an Associate Professor in the Department of Computer Science at University of Missouri-Columbia. His current research interests include distributed and cloud computing, computer networking, and cyber security. He is a Senior Member of IEEE.

Mauro Lemus Alacron received his MS in Mathematics and Computer Science from the McNeese State University. He is currently pursuing his Ph.D. in Computer Science at the University of Missouri- Columbia. His research interests include cloud computing and healthcare data analytics.

Naga Ramya Bhamidipati received her BE degree from Chaitanya Bharathi Institute of Technology, India, in 2017. She is currently pursuing her M.S. degree in Computer Science at University of Missouri-Columbia. Her research interests include cloud computing, blockchain technology, data analytics, and artificial intelligence.

Abu Mosa is the Director of Research Informatics at the University of Missouri (MU) School of Medicine (SOM). Dr. Mosa, also, has an affiliate faculty appointment in the MU Informatics and Data Science Institute and an adjunct appointment in the Electrical Engineering and Computer Science department at MU. He is the lead informatician for building the research informatics infrastructure at the MU SOM. Dr. Mosa's research involves developing the capabilities to implement REDCap, i2b2 (600,000 patients record from the EMR), HealthFacts (a de-identified data set with over 60 million patient records pulled from Cerner clients), SAS 9.4 data analytic server, and secured high performance computing capabilities. He has published more than 16 refereed publications (13 journal articles and 3 full-length conference proceedings) in health/medical informatics. Dr. Mosa is the site-PI for MU's participation in the Greater Plains Collaborative, which is a Clinical Data Research Network (CDRN) funded by the Patient Centered Outcomes Research Institute (PCORI). He is also currently leading three projects funded by Missouri Department of Health and Senior Services and Roche Molecular Systems. He is a co-investigator on NIH, AHRQ and PCORI funded studies for enterprise-wide large scale electronic data capture and healthcare data management studies.

Khaled Salah received the B.S. degree in computer engineering, with a minor in computer science, from Iowa State University, Ames, IA, USA, in 1990, the M.S. degree in computer systems engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 1994, and the Ph.D. degree in computer science from the Illinois Institute of Technology, in 2000. He is currently a Full Professor with the Department of Electrical and Computer Engineering, Khalifa University of Science and Technology, UAE. He has over 220 publications and three U.S. patents, has been giving a number of international keynote speeches, invited talks, tutorials, and research seminars on the subjects of blockchain, IoT, fog and cloud computing, and cybersecurity. He is currently leading a number of projects on how to leverage blockchain for healthcare, 5G networks, combating deepfake videos, supply chain management, and AI. He has served as the Chair for the Track Chair of IEEE Globecom 2018 on Cloud Computing. He is an Associate Editor of IEEE Blockchain Tech Briefs, and a member of the IEEE Blockchain Education Committee.