Maintaining Data Integrity in Electronic Health Records with Hyperledger Fabric

Marten Kask[1], Toomas Klementi[1], Gunnar Piho[2] and Peeter Ross[1]

[1] *TalTech, Department of Health Technologies, Akadeemia Str 15A, Tallinn, 12618 Estonia*
[2] *TalTech, Department of Software Science, Akadeemia Str 15A, Tallinn, 12618 Estonia*

Abstract

Different health data are collected and processed by different medical information systems. Generally, health data is collected by hospitals and stored in the form of Electronic Health Records (EHR). However, as this data contains confidential data, it has become a target for cyber-attacks. Also, as the medical data contains life-critical information, its legitimate origin, reliability and trustworthiness - i.e., data integrity - are particularly important. In the last years, blockchain - a timestamped hierarchical and chronologically-ordered chain of blocks has been proposed as a suitable solution to address the challenges with data integrity. Hyperledger Fabric is an open-source distributed ledger platform allowing enterprise-grade-level solutions to be developed. Therefore, this study aims to present a part of the initial architecture for supporting the integrity of EHRs when exchanging them between organizations. The architecture is based on a blockchain-based system using Hyperledger Fabric technology. Additionally, evaluation methods are proposed to analyze the credibility of architecture in future works.

Keywords

blockchain, Hyperledger Fabric, Electronic Health Record, integrity

1. Introduction

Different health data are collected and processed by various medical information systems. Primarily, health data is collected by hospitals and stored in the form of Electronic Health Records (EHR). EHRs are sets of health data items (observations, measurements, treatments, dietary, etc.) and are usually, among other things, signed to prevent any changes to records after the data entry [1]. EHRs aim to provide efficient availability of accurate data in a diverse clinical setting [2].

Health data has become a target for cyber-attacks as this kind of data is widely collected. It is defined that data integrity measures the sanity of the data, i.e., it originated from a legitimate source [3] and is reliable and trustworthy [4]. A monetary loss can be gained when an attacker has accessed the data, and a ransom is required to restore access [5]. Often, the data in the health information systems is stored in a centralized database which raises the risk that in case of

HEDA 2023: the 3rd International Workshop on Health Data (https://conf.researchr.org/home/staf-2023/heda-2023). Co-located with STAF 2023, 18–21 July, Leicester, United Kingdom.

marten.kask@taltech.ee (M. Kask); toomas.klementi@taltech.ee (T. Klementi); gunnar.piho@taltech.ee (G. Piho); peeter.ross@taltech.ee (P. Ross)

0000-0001-5437-783X (M. Kask); 0000-0002-8260-526X (T. Klementi); 000-0003-4488-338 (G. Piho); 0000-0003-1072-7249 (P. Ross)

unauthorized access, all the stored data can be compromised. These incidents can significantly reduce trust in organizations that store and process medical data but cannot provide and ensure security. Also, preserving data integrity has become a challenge for medical institutions as the organizations and the data they collect are complex. Also, it has been highlighted that maintaining data integrity is a more critical challenge than the other cyber threats [6]. This is because tampered medical information can be life-threatening for patients.

Blockchain is a discovery in secure computing that provides decentralized authority in an open networked system [7]. The central concept of blockchain is to replace the centralized database with authoritative access control. It has been created and maintained as a hierarchical and chronologically-ordered chain of blocks, including timestamps, since its inception in 2009 when Bitcoin was launched.

Blockchain is frequently defined as a distributed ledger [8]. A ledger is a data structure where the transactions are formed into an ordered list, for example, monetary transactions between multiple financial institutions.

Many studies (e.g., [6, 5, 9, 10, 11, 12, 13]), have analyzed the solutions to the issues related to data integrity. They concluded that blockchain-based technology appropriately addresses the health data integrity issues mentioned above. Therefore, this study aims to propose and evaluate a blockchain-based system that is used to support maintaining the integrity of EHRs by using Hyperledger Fabric technology.

2. Technologies in use

2.1. Hyperledger Fabric

Hyperledger Fabric is a permissioned, distributed ledger technology (DLT) platform designed to facilitate the development of enterprise-level blockchain applications [14], [15]. Permissoned blockchains, in contrast to the permissionless blockchains (e.g., BitCoin), allow operating blockchain only among a set of identified and verified participants. In general, that means certain defined rules must be fulfilled before one can join the network, and other members must agree and confirm new participants. This is particularly suitable in the context when transactions related to sensitive data like EHRs. It provides a modular architecture and offers features such as scalability, security, and confidentiality. Industries such as supply chain management, healthcare, financial services, and IoT leverage the benefits of Hyperledger Fabric.

The core components of Fabric's architecture include nodes, channels, chaincode, and the membership service provider (MSP). There are two types of nodes: peer nodes and orderer nodes. Peer nodes can be endorsers, which simulate and endorse transactions, or committers, which validate and commit transactions to the ledger. Orderer nodes establish the order of transactions and create blocks.

Channels are private communication pathways between network members that enable data and transaction isolation. They provide a secure environment for executing chaincode and sharing data between authorized participants.

Chaincode, also known as smart contracts, contains the business logic for processing transactions. It is deployed and executed on the blockchain network and facilitates interactions between participants. The lifecycle of chaincode includes installation, instantiation, and execution.

Hyperledger Fabric uses a pluggable consensus mechanism, allowing flexibility in choosing the most suitable consensus protocol for a particular use case. The default mechanism, called "Raft," is a crash fault-tolerant consensus algorithm that provides high performance and scalability.

The MSP handles identity management and access control within the network. It uses digital certificates and certificate authorities (CAs) to verify and authenticate the identity of participants.

Although Hyperledger Fabric offers various security and privacy features, such as the storage of sensitive data in separate, private ledgers [16]; in this study, it is proposed that the private data itself is not maintained in the Hyperledger Fabric blockchain. Encryption and decryption capabilities protect the confidentiality of data, while access control ensures that only authorized participants can access specific resources. Integration with existing systems is possible through APIs and other interfaces that make the integration with the decentralized storage uncomplicated.

In conclusion, Hyperledger Fabric offers a robust and secure foundation for building enterprise-level blockchain applications. Its modular architecture and customizable features make it a popular choice for various industries looking to leverage blockchain technology and therefore, applicable for proposal in this study.

2.2. Decentralized Storage

Decentralized storage is an emerging serverless technology for securely storing large amounts of data, including sensitive personal data like electronic health records [17]. It does not depend on large servers in huge data centers, but instead, the data is stored in a peer-to-peer network operated by volunteers and containing thousands, potentially millions of nodes. No central authority exists in the network; it is only governed by the protocol implemented in the software running on the nodes.

Data uploaded to such a network is split into multiple small pieces that are then evenly distributed among the nodes, with each piece stored on a different node. To lower the risk of data being lost due to a node leaving the network, each piece is stored not just on a single node but on a set of nodes. Additionally, each piece can be individually encrypted. This means that the network is completely trustless - every node is ignorant of the content and owner of the pieces of data it is storing. Data on such networks is addressed by the hash of its content, which means that its integrity can easily be verified.

The whole data set can be retrieved by its root hash initially known only to the owner of the data. The owner can share the data with third parties by sharing the root hash. It is also conceivable to use other mechanisms for data sharing – by exposing the data through an API etc.

An example of decentralized storage is Ethereum Swarm [18]. It is a distributed storage platform and content delivery network and is designed to function as a foundational infrastructure layer for the wider Ethereum ecosystem. Swarm aims to provide a fully decentralized, redundant, and self-sustaining network for serving and storing data. An illustration of data upload and storage in Swarm [19] is provided in figure 1.

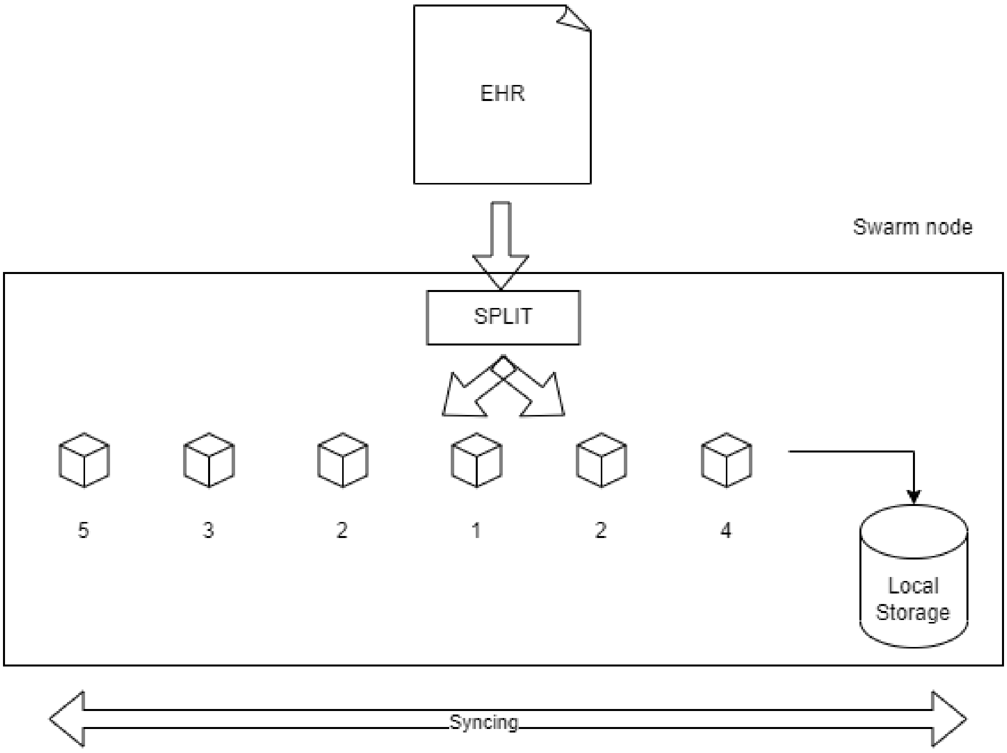

Figure 1: Uploading and storing data in Swarm

3. Proposed Solution

3.1. Current Process of Managing EHRs

Nowadays, health records are managed as EHRs [20]. Managing (including exchanging) them is a crucial process in modern healthcare, facilitating improved patient care, efficient service delivery, and a more coordinated approach to health management. The process usually involves the following steps:

- **Patient Identification**: The first step is to correctly identify the patient whose records need to be exchanged. This can be done using various identifiers like name, date of birth, and unique patient identification numbers.
- **Consent**: Patient consent is a fundamental prerequisite for the exchange of medical records. The healthcare provider must obtain explicit permission from the patient or their representative to share their health records. This typically involves explaining why the records are being shared, with whom, and for what purpose.
- **Record Retrieval**: Once the appropriate permissions are in place, the healthcare provider retrieves the necessary patient records from their electronic health record (EHR) system.
- **Data Format Standardization**: To ensure that the receiving system can read and understand the data, the patient records are converted into a standardized format, such as HL7, FHIR, or CCD, which are common healthcare interoperability standards.

- **Data Transmission**: The records are then transmitted securely (e.g., encrypted) to the intended recipient. This is typically done via a health information exchange (HIE), a secure network designed specifically for the transfer of health information.
- **Record Reception and Integration**: Upon receiving the patient data, the recipient's EHR system imports the records. This includes mapping the incoming data to the appropriate fields within their own system.
- **Verification and Review**: The recipient healthcare provider then verifies the information and reviews the records to ensure accuracy and completeness.
- **Utilization**: Finally, the healthcare provider uses the records to make informed decisions about patient care.

Data security and patient privacy are paramount throughout this process. This is often ensured by using encryption, secure networks, and strict access controls. Laws and regulations, such as HIPAA in the U.S and GDPR in EU, also mandate specific data handling practices to protect patient information during this process.

In this study, the focus is on the two simplified scenarios related to the EHR exchange. The first is the use case, where the doctor logs in to their hospital information system by providing credentials and entering the patient's health data. In a generalized approach, the data is stored in a centralized database that is accessible by other hospitals in the region. The other use case is that a patient wishes to access their health records. The patient logs in to the patient web portal that is provided by another healthcare provider in the region and opens a dedicated page where the health data is displayed.

These use cases rely significantly on internal business processes that other health organizations cannot verify transparently. For example, hospitals can agree on which authentication methods and access are required for the doctors to be allowed to enter health data in the database (and patients to request their health data). However, the other organizations must trust that other parties follow the agreed rules. External audits are often carried out to verify that agreements are followed, but instant verifications for every transaction are usually impossible. The architecture described in the following sections addresses the described issues and proposes solutions for how transparency can be increased. Additionally, the study addresses the concern related to health data storage in a centralized database, whereas a decentralized alternative is provided instead.

3.2. Deployment

According to the description provided in the previous sections, the following deployment setup was constructed. It is expected that at least two different organizations (e.g., hospitals) exchange EHRs. A descriptive diagram is provided in figure 2.

In accordance with Hyperledger Fabric requirements, the blockchain network is initiated by setting up the channel. Channel consists of peers that store the copy of distributed ledger. Each organization is represented by the peer. There is a channel configuration that defines how the peers are added to the channel. Smart contracts that are described in the following section are invoked by the client applications that are represented by the hospital information systems (Hospital IS) in the deployment described in this study.

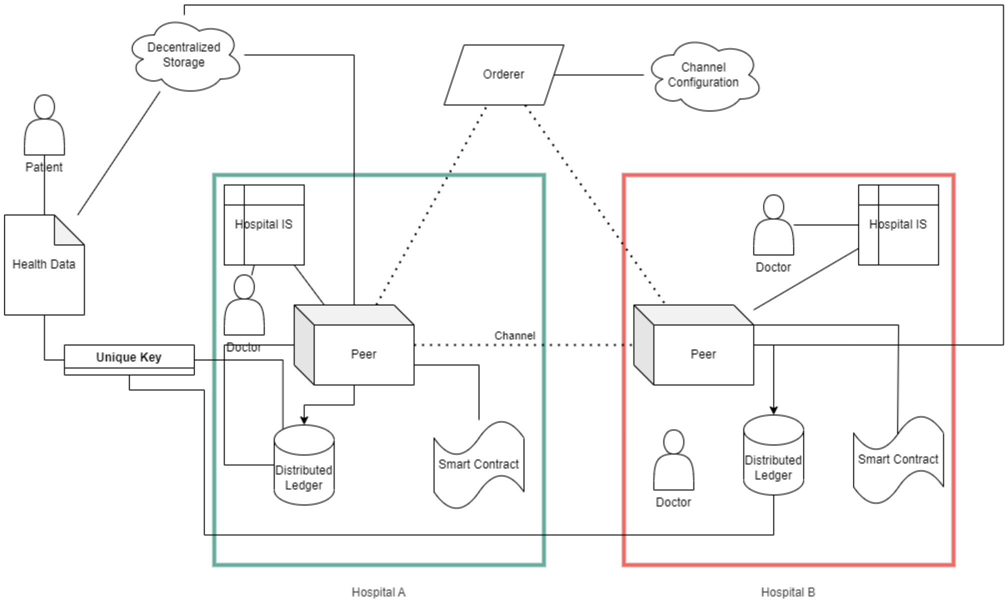

Figure 2: Deployment for the Blockchain-based System

3.3. Smart Contracts

In blockchain, Smart contracts generally define the rules that allow creating, reading, updating and deleting business objects in the ledger [21]. They define a transaction logic that allows controlling the lifecycle of these business objects. In Hyperledger Fabric, it is developed in chaincode, and all the members have to agree to have them implemented. Smart contracts are invoked by the client applications that are considered as hospital information systems in this paper.

The following main smart contracts are proposed in this study:

- **SC1** - authorizing the Doctor
  - **Description**: It is expected that only authorized persons are allowed to enter new Electronic Health records. Thus, the smart contract shall verify:
    - a) the person who enters the EHR data (e.g., using technology that corresponds to the EU regulation of eIDAS (electronic IDentification, Authentication and trust Services);
    - b) the verified person is registered in the acceptable registry (e.g., Health Care Workers Registry).

As described in the beginning of this section, all the organizations have to agree to the smart contracts. That means each organization recognizes the agreed authorization methods and registers of health workers.

- **SC2** - authorizing the Patient

- **Description**: It is expected that only authorized persons are allowed to view/retrieve their Electronic Health records. Thus, the smart contract shall:
  - a) verify the person who views/retrieves the EHR data (e.g., using technology that corresponds to the EU regulation of eIDAS (electronic IDentification, Authentication and trust Services);
  - b) decrypt the unique hash of the EHR according to the authentication data

Additionally, smart contracts for ensuring data quality can be introduced to verify, for example, that contradictory information is not entered but it is not in the scope of the present study.

3.4. Hyperladger Fabric with Decentralized Storage

Although Hyperledger Fabric and decentralized storage (e.g., Ethereum Swarm) are both described with a keyword as "decentralized", yet they mostly serve different purposes in the realm of decentralized systems.

Hyperledger Fabric is a blockchain framework implementation intended for developing applications or solutions with a modular architecture. It allows components, such as consensus and membership services, to be plug-and-play and offers features like channels for private communications between a specific set of members. It is primarily used for developing private, permissioned blockchain networks where all participants are known and identifiable.

On the other hand, decentralized storage like Ethereum Swarm is a distributed storage platform and content delivery service which allows a network of peers to store and distribute chunks of arbitrary data. It is designed to function as a foundational layer of infrastructure for the Ethereum ecosystem, providing a fully decentralized and resilient way of storing and serving digital content.

Therefore, we propose that these two systems complement each other. Blockchains like Hyperledger Fabric are not designed for storing large amounts of data as it would quickly become impractical due to scalability issues and the necessity for all nodes to carry a full copy of the blockchain. Instead, we propose that the actual data is stored on decentralized storage, and then store the reference to that data on the Hyperledger Fabric blockchain. This way, it is possible to get the data integrity, audibility, and traceability benefits of a blockchain, and the scalable, resilient data storage capabilities from decentralized storage.

3.5. Proposed Process for Entering a New EHR

In this subsection, a simplified proposed process for entering a new EHR is described. To enter a new Electronic Health Record, Doctor uses a hospital information system. After entering the health data of the patient, the system constructs an EHR according to the standard and stores it temporarily in the information system. Based on the constructed EHR, a unique hash of the record is generated using a defined algorithm. After that, the generated hash of the EHR is encrypted and submitted to be stored in Hyperledger Fabric blockchain.

After the request to submit the hash in the blockchain, Hospital A Peer creates a smart contract request (SC1), validates, signs and invokes it. In addition, the request is sent to the peers of other hospitals as well who all are following the same procedure. If the request is

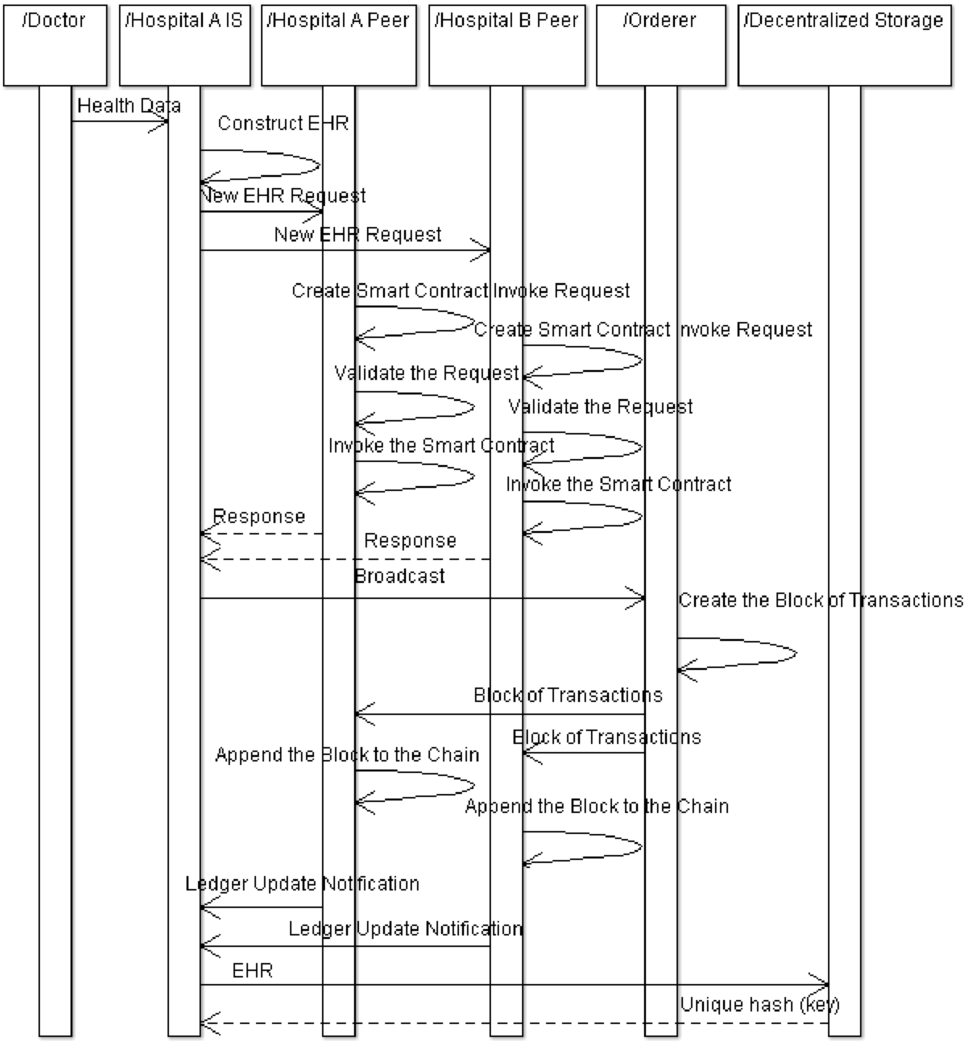
/Doctor
/Hospital A IS
/Hospital A Peer
/Hospital B Peer
/Orderer
/Decentralized Storage
Health Data
Construct EHR
New EHR Request
New EHR Request
Create Smart Contract Invoke Request
Create Smart Contract Invoke Request
Validate the Request
Validate the Request
Invoke the Smart Contract
Invoke the Smart Contract
Response
Response
Broadcast
Create the Block of Transactions
Block of Transactions
Block of Transactions
Append the Block to the Chain
Append the Block to the Chain
Ledger Update Notification
Ledger Update Notification
EHR
Unique hash (key)

Figure 3: Sequence Diagram of the Proposed Process of Entering a New EHR

successful, all peers send a response to the Hospital A IS. Subsequently, Hospital IS validates the responses and broadcasts them to the orderer.

Orderer receives the responses and orders them chronologically. After that, the block of transactions is generated and delivered to all the peers. Then the peers validate and append the block to the chain, and send a notification to the Hospital A IS that the ledger is updated. If Hospital A IS receives the confirmation that the ledger has been updated, the constructed EHR with the unique hash is sent to the decentralized storage and deleted from the Hospital A IS's temporary storage. The process is illustrated in a figure 3.

3.6. Proposed Process for Retrieving a EHR

In this subsection, a proposed process for retrieving an EHR is described. It follows a similar process as entering a new EHR as described in the previous subsection. To simplify the example, it is expected that Hospital A's information system provides a Patient Portal. However, on the broader picture it can be assumed that special organizations exist in the blockchain channel that are focused on providing user interfaces for patients.

To retrieve an existing Electronic Health Record, the Patient uses Patient Portal. After providing identification data, the system submits a request to Hyperledger Fabric blockchain to decrypt and retrieve a unique hash that can be used to access EHR from decentralized storage. After the request to retrieve a unique hash is submitted in blockchain, Hospital A Peer creates a

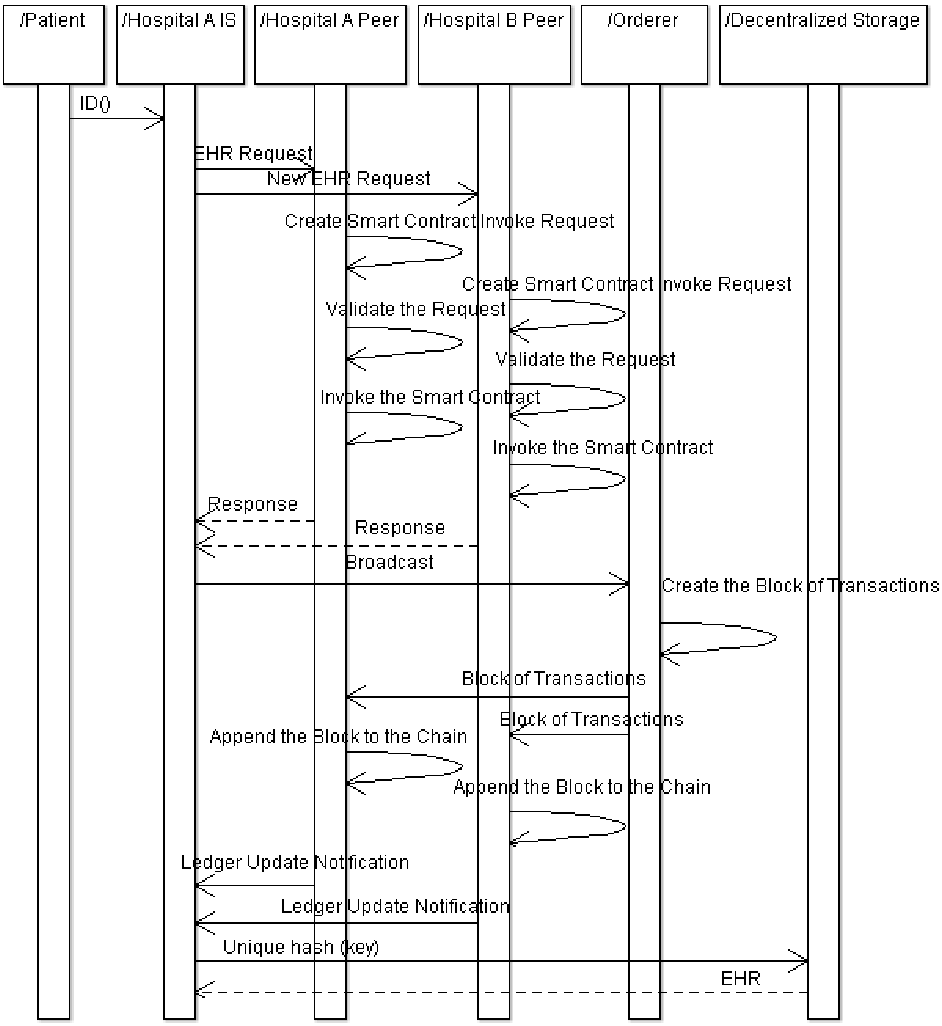
/Patient
/Hospital A IS
/Hospital A Peer
/Hospital B Peer
/Orderer
/Decentralized Storage
ID()
EHR Request
New EHR Request
Create Smart Contract Invoke Request
Create Smart Contract Invoke Request
Validate the Request
Validate the Request
Invoke the Smart Contract
Invoke the Smart Contract
Response
Response
Broadcast
Create the Block of Transactions
Block of Transactions
Block of Transactions
Append the Block to the Chain
Append the Block to the Chain
Ledger Update Notification
Ledger Update Notification
Unique hash (key)
EHR

Figure 4: Sequence Diagram of the Proposed Process of Entering a New EHR

smart contract request (SC2), validates, signs, and invokes it. In addition, the request is sent to the peers of other hospitals as well who all are following the same procedure. If the request is successful, all peers send a response to the Hospital A IS. Subsequently, Hospital IS validates the responses and broadcasts them to the orderer.

Orderer receives the responses and orders them chronologically. After that, the block of transactions is generated and delivered to all the peers. Then the peers validate and append the block to the chain, and send a notification to the Hospital A IS that the ledger is updated with a new entry. If Hospital A IS receiving confirmation that the ledger has been updated, the decrypted unique hash is provided to the Patient Portal that uses it to retrieve the EHR from the decentralized storage. The process is illustrated in figure 4.

4. Discussion

4.1. Related Works

Several studies have been conducted to improve the data integrity of EHRs by implementing blockchain technology. Pilares et al. [22] have proposed an EHRChain framework that is enabled by dual-blockchains based on Hyperledger Sawtooth and InterPlanetary File System (IPFS). They have outlined that EHR data cannot be tampered with when the supermajority of nodes are trustworthy. Kim et al. have [23] proposed a secure protocol for cloud-assisted EHR system using Hyperledger Fabric blockchain that log transactions are used to provide data integrity and access control. Another blockchain based application is introduced by Faroug et al. [24] using Hyperledger Fabric and Ethereum platforms. A Hyperledger Fabric-based solution is proposed by Kumar and Dakshayini [25].

Additionally, BiiMED (a Blockchain framework for Enhancing Data Interoperability and Integrity) has been proposed by Jabbar et al. [26] regarding EHR-sharing. However, in addition to the blockchain's properties, Decentralized Trusted Third Party Auditor (TTPA) is used for ensuring data integrity as well. After retrieving shared data from another medical facility, the Health Information System will compare the hash of the received data with the stored hash to verify the integrity of the received data. Kalaipriya et al. [27] describe a blockchain-based design using Ethereum framework for current Electronic Health Records (EHR) frameworks where two smart agreements, classified contracts, and user record associated contracts are introduced as well.

Ghayvat et al. [28] propose a scheme that integrates blockchain (BC)-based confidentiality-privacy (CP) preserving scheme, CP-BDHCA, that operates in two phases. It is highlighted that blockchain addresses the healthcare cloud and application limitations about agreements, accountability, rights management, and data integrity in healthcare big data. A more general approach is described by Yuan et al. [29]. A scheme that integrates the Ethernum blockchain and third-party auditors (TPA) is studied by Liu et al. [30]. In addition to the blockchain attributes, user can send an auditing request for data integrity. The TPA examines the integrity of health records through the auditing proof upon receipt. A system design integrating edge computing paradigm, blockchain technology and Inter-Planetary File System (IPFS) is another study by Makina et al. [31] where in addition to blockchain's attributes, supplementary measures are introduced to ensure data integrity. MB-EHR (Multilayer Blockchain-based EHR) proposes a layered blockchain structure to support better the operational hierarchy in health organisations where PDP-like data verification mechanism protects data integrity [32]. Furthermore, authors [33] present a MEChain, a multilayer blockchain structure, which aims to solve the adoption, storage and consensus problems when implementing blockchain in EHR systems. They outline that data synchronisation method provides data verification and retrieve mechanisms to protect data integrity.

Ajayi et al. [34] propose a blockchain-based solution that facilitates a scalable and secured inter-healthcare EHRs exchange where the integrity and consistency of EHR requests and replies is verified and presented in a standard format to make them easily understandable for different healthcare systems. Shahnaz et al. [35] propose a framework that implements blockchain technology for EHR and secondly to provides secure storage of electronic records by defining granular access rules for the users of the proposed framework where access rules ensure that patients' private data or medical records are not accessible and remain temper-proof.

A wrapper layer integration mechanism, named as the blockchain handshake, between the existing cloud-based EHR management system and public blockchain network to develop a tamper-proof health record management system is introduced by Rahman et al. [36]. Akbar et al. [37] propose an Ethereum blockchain-based solution where a function to ensure that the content of the smart contract from the system does not change by comparing the bytecode of the deployed smart contract with the one that has not deployed.

A Scrybe, a permissioned blockchain, to store proof of clinical trial data provenance is introduced by Oakley et al. [38]. They illustrate how Scrybe addresses each control and the limitations of the Ethereum-based blockchains. Khan et al. [39] propose a blockchain-based framework to facilitate health data availability and sharing. An internet-inspired framework (ChainNet) to facilitate interoperability within blockchain-based systems whereby two systems within independent Blockchain networks can securely exchange data with each other is introduced by Abdullah et al. [40].

All the included studies have addressed maintaining data integrity by outlining the general blockchain attributes, i.e., that the records are immutable. What is more, some authors have described maintaining data integrity in more detail. e.g., used cryptography. Also, some authors introduced additional tools like third-party auditors in addition to the general attributes of blockchain to maintain the data integrity. Many studies propose that medical data itself is stored in blockchain. This, however can negatively influence the scalability as the amount of data can be high and privacy issues as all the blockchain members are able to read the data from ledger.

Therefore, the solution proposed in this study introduces an integrated solution where Hyperledger Fabric is used to control the business processes related to EHR management and decentralized storage to store medical information in a distributed form. To sum up, there are several differences between the existing works and advantages that the study proposes. For example, to highlight the most important ones, this study proposes integrating eIDAS and acceptable registries to increase trust in authentication and authorization. Furthermore, some studies presented solutions where the health data is stored on a blockchain, but this brings scalability problems and means that all the nodes somehow store a copy of the data.

Although there are studies that suggest that only a hash of the data instead of the whole record is stored on a blockchain, there needs to be more emphasis on storing the actual data. This research proposes complementary integration with decentralized storage, ensuring scalable and resilient data storage. In ransom attacks, the intruder has often accessed the whole centralized database and encrypted it. Nevertheless, suppose an attacker obtains access to one of the decentralized nodes. In that case, it is only possible to receive a fragment of the data with which it is impossible to assemble the entire record.

4.2. Future Work: A Possible Evaluation Methodology for the Proposed Solution

To verify whether the proposed solution resolves the previously outlined concerns and does not create new ones, an evaluation methodology shall be addressed in the future works. It shall be comprehensive and focused on critical aspects like security, privacy, interoperability, scalability, and performance. It also should be adapted to the specific use cases and organizations. Engaging stakeholders, including healthcare providers, IT experts, and patients, in the evaluation process will ensure that the chosen solution architecture addresses the needs and concerns of all parties involved.

We have planned a possible evaluation methodology as follows:

1. **Define Evaluation Criteria**: Establish clear and quantifiable criteria to assess the proposed solution architecture. These criteria may include:
   - Data Integrity
   - Data Confidentiality
   - Access Control
   - Interoperability
   - Scalability
   - Performance
   - Compliance with standards and regulations
   - User Experience
2. **Develop Evaluation Metrics**: For each criterion, define specific metrics that can be used to objectively measure the solution's performance. Examples include:
   - Data Integrity: percentage of successful data validation checks, percentage of data discrepancies
   - Data Confidentiality: encryption strength, percentage of unauthorized data breaches
   - Access Control: number of successful/failed authorization attempts, time to grant/revoke access
   - Interoperability: number of successfully integrated systems, data exchange success rate
   - Scalability: response time under varying loads, number of concurrent users supported
   - Performance: transaction processing time, system response time
   - Compliance: number of compliance checks passed, audit results
   - User Experience: user satisfaction ratings, time to complete tasks

3. **Create Test Scenarios and Benchmarks**: Design test scenarios and benchmarks that simulate real-world use cases and challenges. These tests should be conducted in a controlled environment to compare different solution architectures fairly.
4. **Implement and Test Solution Architectures**: Build prototypes or proof-of-concept implementations to evaluate each solution architecture. Conduct tests using the designed scenarios and benchmarks, and collect data on the defined metrics.
5. **Analyze Results**: Analyze the collected data and compare the performance of each solution architecture against the evaluation criteria and metrics. Identify strengths and weaknesses in each architecture and determine areas for improvement.
6. **Rank and Select the Best Solution**: Based on the analysis, rank the solution architectures and select the one that best meets the evaluation criteria. Consider trade-offs and the overall alignment of the solution with the goals and requirements of the EHR system.
7. **Iterate and Refine**: Continuously iterate on the chosen solution architecture, refining and optimizing it to enhance its performance and meet evolving needs.

Due to the huge amount of data that is exchanged in the medical domain, the scalability issue can be the first one to address when improving the proposal and developing a Proof-of-Concept. For example, it may not be reasonable that all the participating members in the Hyperledger Fabric network maintain the whole history of ledger all the time. Generally, in the everyday use cases, it might be sufficient for the doctor or patient to be sure that the EHR they currently process is the most recent and valid. The fundamental solution to approach that can be by introducing blockchain sharding - splitting the blockchain into shards or partitions that allow processing more transactions in parallel.

Another challenge to address is future-proofing. As the computational power is growing over time and the distribution of quantum computers can put the cryptography used in Hyperledger Fabric and decentralized storage on the spot. Thus, the solutions should take into account that cryptographic providers could be upgraded over time. On the other hand, the introduction of quantum computers can be the solution to the scalability concerns outlined previously.

Additionally, the membership service provider that is used in Hyperledger Fabric to prove the identity of blockchain participants shall be analyzed in the future as it was not in the scope of this paper. For example, the possibility of integration of European Digital Identity can be examined and determined.

5. Conclusion

This paper introduced an initial architecture concept for supporting EHR integrity when exchanging them between organizations. The architecture is based on a blockchain-based system using Hyperledger Fabric technology, while the EHR data is stored in decentralized storage. Hyperledger Fabric and decentralized storage concepts are introduced, and their main components are described.

To present a proposed solution, a current process of managing EHRs are briefly outlined. Based on the concepts of introduced technologies in use and the current process, a deployment and example use cases are represented.

In the following, a short overview of related works on improving data integrity of EHRs is provided. It is outlined that all the studies address that data integrity can be maintained by the main attributes of blockchain. Also, many studies described that health data itself is also stored in blockchain. As this can have negative implications, this study proposes an integrated solution where Hyperledger Fabric is used to control the business processes related to EHR management and decentralized storage to store medical information in a distributed form.

Additionally, evaluation methods are proposed to analyze the credibility of architecture in future works. Challenges to address, like future-proofing and membership service provider, are also outlined to be addressed in prospective studies.

Acknowledgments

This work in the project 'ICT programme' was supported by the European Union through the European Social Fund.

References

[1] T. Kanwal, A. Anjum, A. Khan, Privacy preservation in e-health cloud: taxonomy, privacy requirements, feasibility analysis, and opportunities, Cluster Computing 24 (2021) 293–317.

[2] P. Vimalachandran, H. Wang, Y. Zhang, B. Heyward, F. Whittaker, Ensuring data integrity in electronic health records: a quality health care implication, in: 2016 International Conference on Orange Technologies (ICOT), IEEE, 2016, pp. 20–27.

[3] V. G. Garagad, N. C. Iyer, H. G. Wali, Data integrity: a security threat for internet of things and cyber-physical systems, in: 2020 International Conference on Computational Performance Evaluation (ComPE), IEEE, 2020, pp. 244–249.

[4] S. Manu, G. Bhaskar, Securing sensitive data in body area sensor network using blockchain technique, in: 2020 5th International Conference on Communication and Electronics Systems (ICCES), IEEE, 2020, pp. 1–5.

[5] A. K. Pandey, A. I. Khan, Y. B. Abushark, M. M. Alam, A. Agrawal, R. Kumar, R. A. Khan, Key issues in healthcare data integrity: Analysis and recommendations, IEEE Access 8 (2020) 40612–40628.

[6] M. Zarour, M. Alenezi, M. T. J. Ansari, A. K. Pandey, M. Ahmad, A. Agrawal, R. Kumar, R. A. Khan, Ensuring data integrity of healthcare information in the era of digital health, Healthcare Technology Letters 8 (2021) 66–77.

[7] R. Zhang, R. Xue, L. Liu, Security and privacy on blockchain, ACM Computing Surveys (CSUR) 52 (2019) 1–34.

[8] M. Abdelhamid, G. Hassan, Blockchain and smart contracts, in: Proceedings of the 8th International Conference on Software and Information Engineering, 2019, pp. 91–95.

[9] A. Hasselgren, K. Kralevska, D. Gligoroski, S. A. Pedersen, A. Faxvaag, Blockchain in healthcare and health sciences—a scoping review, International Journal of Medical Informatics 134 (2020) 104040.

[10] M. Kask, G. Piho, P. Ross, Systematic literature review of methods for maintaining data integrity, in: Advances in Model and Data Engineering in the Digitalization Era: MEDI

2021 International Workshops: DETECT, SIAS, CSMML, BIOC, HEDA, Tallinn, Estonia, June 21–23, 2021, Proceedings 10, Springer, 2021, pp. 259–268.

[11] A. Gonzales, S. R. Smith, P. Dullabh, L. Hovey, K. Heaney-Huls, M. Robichaud, R. Boodoo, Potential uses of blockchain technology for outcomes research on opioids, JMIR Medical Informatics 9 (2021) e16293.

[12] K. Fan, S. Wang, Y. Ren, H. Li, Y. Yang, Medblock: Efficient and secure medical data sharing via blockchain, Journal of medical systems 42 (2018) 1–11.

[13] M. Jones, M. Johnson, M. Shervey, J. T. Dudley, N. Zimmerman, Privacy-preserving methods for feature engineering using blockchain: review, evaluation, and proof of concept, Journal of medical Internet research 21 (2019) e13600.

[14] Hyperledger fabric whitepaper, https://www.hyperledger.org/wp-content/uploads/2020/03/hyperledger_fabric_whitepaper.pdf, note = Accessed: 2023-05-20, ????.

[15] Introduction - hyperledger-fabricdocs main documentation, https://www.hyperledger.org/wp-content/uploads/2020/03/hyperledger_fabric_whitepaper.pdf, ????. Accessed: 2023-05-20.

[16] Private data - hyperledger-fabricdocs main documentation, https://hyperledger-fabric.readthedocs.io/en/latest/private-data-arch.html, ????. Accessed: 2023-05-20.

[17] T. Klementi, K. J. I. Kankainen, G. Piho, P. Ross, Prospective research topics towards preserving electronic health records in decentralised content-addressable storage networks 3264 (2022). URL: https://ceur-ws.org/Vol-3264/HEDA22_paper_7.pdf.

[18] Welcome! | swarm bee client, https://docs.ethswarm.org/docs/, note = Accessed: 2023-05-20, ????

[19] P. Febrero, An overview of ethereum swarm: A decentralised filestore, 2020. URL: https://finance.yahoo.com/news/overview-ethereum-swarm-decentralised-filestore-140009996.html.

[20] A. Shibu, A. M, A. T. Anilkumar, A. Radhakrishnan, S. Izudheen, Secure storage and retrieval of electronic health records, in: 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), 2022, pp. 1–5. doi:10.1109/IC3SIS54991.2022.9885484.

[21] Smart contracts and chaincode -; hyperledger-fabricdocs main documentation, https://hyperledger-fabric.readthedocs.io/en/release-2.5/smartcontract/smartcontract.html, ???? Accessed: 2023-05-18.

[22] I. C. A. Pilares, S. Azam, S. Akbulut, M. Jonkman, B. Shanmugam, Addressing the challenges of electronic health records using blockchain and ipfs, Sensors 22 (2022). URL: https://www.mdpi.com/1424-8220/22/11/4032. doi:10.3390/s22114032.

[23] M. Kim, S. Yu, J. Lee, Y. Park, Y. Park, Design of secure protocol for cloud-assisted electronic health record system using blockchain, Sensors 20 (2020). URL: https://www.mdpi.com/1424-8220/20/10/2913. doi:10.3390/s20102913.

[24] A. Faroug, M. Demirci, Blockchain-based solutions for effective and secure management of electronic health records, in: 2021 International Conference on Information Security and Cryptology (ISCTURKEY), 2021, pp. 132–137. doi:10.1109/ISCTURKEY53027.2021.9654325.

[25] N. Kumar S., M. Dakshayini, Secure sharing of health data using hyperledger fabric based on blockchain technology, in: 2020 International Conference on Mainstreaming Block Chain

Implementation (ICOMBI), 2020, pp. 1–5. doi:10.23919/ICOMBI48604.2020.9203442.

[26] R. Jabbar, N. Fetais, M. Krichen, K. Barkaoui, Blockchain technology for healthcare: Enhancing shared electronic health record interoperability and integrity, in: 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 310–317. doi:10.1109/ICIoT48696.2020.9089570.

[27] R. Kalaipriya, S. Devadharshini, R. Rajmohan, M. Pavithra, T. Ananthkumar, Certain investigations on leveraging blockchain technology for developing electronic health records, in: 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1–5. doi:10.1109/ICSCAN49426.2020.9262391.

[28] H. Ghayvat, S. Pandya, P. Bhattacharya, M. Zuhair, M. Rashid, S. Hakak, K. Dev, Cp-bdhca: Blockchain-based confidentiality-privacy preserving big data scheme for healthcare clouds and applications, IEEE Journal of Biomedical and Health Informatics 26 (2022) 1937–1948. doi:10.1109/JBHI.2021.3097237.

[29] L. Q. Yuan, M. E. Rana, Q. A. Maatouk, Enhancing medical data transparency and integrity with blockchain based implementation, in: 2021 Third International Sustainability and Resilience Conference: Climate Change, 2021, pp. 279–285. doi:10.1109/IEEECONF53624.2021.9668137.

[30] X. Liu, Y. Luo, X. Yang, L. Wang, X. Zhang, Lattice-based proxy-oriented public auditing scheme for electronic health record in cloud-assisted wbans, IEEE Systems Journal 16 (2022) 2968–2978. doi:10.1109/JSYST.2021.3138861.

[31] H. Makina, A. B. Letaifa, A. Rachedi, Leveraging edge computing, blockchain and ipfs for addressing ehealth records challenges, in: 2022 15th International Conference on Security of Information and Networks (SIN), 2022, pp. 01–04. doi:10.1109/SIN56466.2022.9970495.

[32] H. Wu, L. Li, H.-y. Paik, S. S. Kanhere, Mb-ehr: A multilayer blockchain-based ehr, in: 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), 2021, pp. 1–3. doi:10.1109/ICBC51069.2021.9461075.

[33] H. Y. Wu, L. J. Li, H.-Y. Paik, S. S. Kanhere, Mechain: A multi-layer blockchain structure with hierarchical consensus for secure ehr system, in: 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2021, pp. 976–987. doi:10.1109/TrustCom53373.2021.00136.

[34] O. Ajayi, M. Abouali, T. Saadawi, Secure architecture for inter-healthcare electronic health records exchange, in: 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2020, pp. 1–6. doi:10.1109/IEMTRONICS51293.2020.9216336.

[35] A. Shahnaz, U. Qamar, A. Khalid, Using blockchain for electronic health records, IEEE Access 7 (2019) 147782–147795. doi:10.1109/ACCESS.2019.2946373.

[36] M. S. Rahman, I. Khalil, P. C. Mahawaga Arachchige, A. Bouras, X. Yi, A novel architecture for tamper proof electronic health record management system using blockchain wrapper, in: Proceedings of the 2019 ACM International Symposium on Blockchain and Secure Critical Infrastructure, BSCI '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 97–105. URL: https://doi.org/10.1145/3327960.3332392. doi:10.1145/3327960.3332392.

[37] I. M. Akbar, A. Bhawiyuga, R. Siregar, An ethereum blockchain based electronic health record system for inter-hospital secure data sharing, in: Proceedings of the 6th International Conference on Sustainable Information Engineering and Technology, SIET

'21, Association for Computing Machinery, New York, NY, USA, 2021, p. 226–230. URL: https://doi.org/10.1145/3479645.3479699. doi:10.1145/3479645.3479699.

[38] J. Oakley, C. Worley, L. Yu, R. R. Brooks, u. Özçelik, A. Skjellum, J. S. Obeid, Scrybe: A secure audit trail for clinical trial data fusion, Digital Threats (2022). URL: https://doi.org/10.1145/3491258. doi:10.1145/3491258, just Accepted.

[39] A. Khan, A. Anjum, Blockchain-based distributed platform for accountable medical data sharing, in: Proceedings of the 14th IEEE/ACM International Conference on Utility and Cloud Computing Companion, UCC '21, Association for Computing Machinery, New York, NY, USA, 2022, pp. 1–8. URL: https://doi.org/10.1145/3492323.3503506. doi:10.1145/3492323.3503506.

[40] S. Abdullah, J. Arshad, M. Alsadi, Chain-net: An internet-inspired framework for interoperable blockchains, Distrib. Ledger Technol. 1 (2022). URL: https://doi.org/10.1145/3554761. doi:10.1145/3554761.