A Privacy-Preserving Healthcare Framework Using Hyperledger Fabric

Charalampos Stamatellis [1], **Pavlos Papadopoulos** [1,*], **Nikolaos Pitropakis** [1,2],
Sokratis Katsikas [3,*] and **William J. Buchanan** [1]

[1] Blockpass ID Lab, School of Computing Edinburgh Napier University, Edinburgh EH10 5DT, UK;
    40333763@live.napier.ac.uk (C.S.); N.Pitropakis@napier.ac.uk (N.P.); B.Buchanan@napier.ac.uk (W.J.B.)
[2] Eight Bells LTD, Nicosia 2002, Cyprus
[3] Department of Information Security and Communication Technology, Norwegian University of Science and
    Technology, 2815 Gjøvik, Norway
[*] Correspondence: pavlos.papadopoulos@napier.ac.uk (P.P.); sokratis.katsikas@ntnu.no (S.K.)

check for updates

Abstract: Electronic health record (EHR) management systems require the adoption of effective technologies when health information is being exchanged. Current management approaches often face risks that may expose medical record storage solutions to common security attack vectors. However, healthcare-oriented blockchain solutions can provide a decentralized, anonymous and secure EHR handling approach. This paper presents PREHEALTH, a privacy-preserving EHR management solution that uses distributed ledger technology and an Identity Mixer (Idemix). The paper describes a proof-of-concept implementation that uses the Hyperledger Fabric's permissioned blockchain framework. The proposed solution is able to store patient records effectively whilst providing anonymity and unlinkability. Experimental performance evaluation results demonstrate the scheme's efficiency and feasibility for real-world scale deployment.

Keywords: distributed ledger technology; electronic health records management; privacy-preserving; healthcare; Hyperledger Fabric; blockchain

1. Introduction

Traditional healthcare practices have often been influenced by the usage of paper-based medical records and these have evolved into electronic patient records [1]. Thus, electronic health records (EHRs) often contain highly sensitive medical information, which is shared among healthcare providers, pharmacies and patients [2]. Current EHR management approaches include distributed or cloud server data storage; this practice can lead to diverse functionalities and financial complications. Further, according to Coventry and Branley [3], healthcare organizations are vulnerable to a number of cybersecurity threats, including malware and ransomware.

Successful compromise of EHR management and storage providers can be achieved by deploying ransomware, which encrypts data until a ransom is paid, or by hacking the EHR repository [3]. The WannaCry ransomware cryptoworm, used in the WannaCry cyber attack that resulted in a loss of £92 million, infected vulnerable servers and computers of the National Health Service (NHS) and encrypted medical data at 80 out of 236 NHS foundations [4]. Further, the Medjack (Medical Device Hijack) cyber attack used malware that infected multiple unprotected medical devices, compromised network defences, breached user anonymity and accessed medical records [3]. Healthcare information is often substantially more valuable than other data sources for trading on the black market: the average cost of a hijacked medical record is approximately $380; this is twice as large as the average cost across all industry-related data breaches [3]. Incidents such as the above demonstrate that the healthcare

sector can suffer substantial financial loss as a result of common attack vectors against traditional EHR databases [5].

With the increasing drive towards EHR, there is a certain urgency for a scalable, immutable, transparent and secure solution to be implemented in order to address the aforementioned challenges. A state-of-the-art approach, invulnerable to common attack vectors, could handle medical data in a decentralized manner in order to avoid the possibility of a single point-of-failure or a single point-of-attack. A key driver is to achieve general agreement amongst the co-operating healthcare providers in order for weak security and insider threats to be prevented or properly responded to.

Current EHR storage and distribution approaches utilize certain solutions in the context of challenging the aforementioned security threats. In particular, they use access control policies such as role-based access control (RBAC) and attribute-based access control (ABAC), which restrict system access to unauthorized users according to specific preassigned roles and attributes [6]. Moreover, a secure and anonymous three-factor authentication protocol could be deployed in the context of privacy-preserving healthcare-oriented wireless sensor networks (WSNs) [7]. Additionally, encryption combined with pseudonymization techniques is deployed to conceal users' identities and preserve anonymity. However, access control policies may leak private information and encryption mechanisms could potentially affect system performance [8].

Contributions

In this paper we propose Privacy-Preserving Healthcare (PREHEALTH), a privacy-preserving EHR management solution, illustrated in Figure 1. PREHEALTH can be used by patients or other healthcare actors that wish to store health records in an immutable distributed ledger that ensures data privacy. This immutable ledger protects data against misuses and abuses, whilst allowing the querying of non-personal data by authorized blockchain participants, for example medical professionals. Additionally, PREHEALTH supports secure auditing, preservation of the privacy of accumulated personal data and isolates data related to auditing and compliance checks.
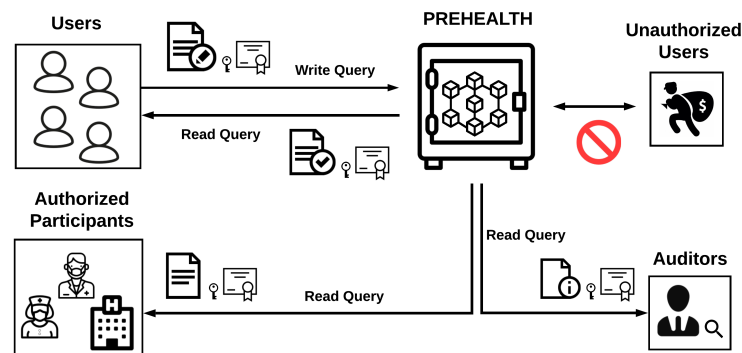

Figure 1. Privacy-Preserving Healthcare overview.

The contributions of our work can be summarized as follows:

- The development of PREHEALTH, a privacy-preserving healthcare data solution. This uses distributed ledger technology and the Identity Mixer (Idemix) suite. Idemix is a zero-knowledge proof (ZKP) cryptographic protocol that provides privacy-preserving features, such as anonymity and unlinkability.
- An evaluation of the robustness and security of PREHEALTH against popular attack vectors.
- Empirical comparison of the performance of PREHEALTH against relevant blockchain solutions in the literature, and against a traditional database that offers related in-context column-level encryption.

The Idemix suite is still under development and currently has certain limitations that would need to be improved in future releases, as the utilization of the aforementioned state-of-art architecture within a health record distribution approach is an essential contribution. With Idemix, a client acting as a blockchain participant may read, update or share private medical records while preserving their anonymity and privacy.

The remainder of the paper is organized as follows. Section 2 provides a background on the adopted distributed ledger technology that is essential to assure the self-sustainability of the paper. The related literature is discussed in Section 3. In Section 4, the proof-of-concept implementation of PREHEALTH is presented, which also serves as our evaluation testbed. Section 5 presents the process and the results of the evaluation of PREHEALTH, and Section 6 discusses our findings. Finally, Section 7 summarizes our conclusions and suggests items for future work.

2. Background

Blockchain is one of the most disruptive technologies of recent years [9]. The term is used to describe a decentralized peer-to-peer network, where an append-only and auditable ledger is updated by non-trusting entities [10]. The number of both validators and end-users of the platform can be specified (permissioned); nevertheless, the public and anonymous (permissionless) access option is also possible.

A permissionless blockchain network on the Ethereum blockchain framework has been proposed to implement a digital couponing system. Efficient non-repudiation and decentralization have been achieved, but scalability and anonymity lag behind [11]. Longo, Podda and Saia [12] examined Subchains, a contemporary blockchain technique, developed on the Bitcoin public ledger, which allows third-party services to store data in their database and synchronizes them with the Bitcoin blockchain network. However, both synchronization and security considerations affect the anonymity and scalability of the proposed scheme.

On the other hand, an example of a permissioned ledger is Hyperledger Fabric, which is an open-source blockchain platform that endorses strong security and identity features [13]. Hyperledger Fabric introduces a novel architecture [14,15] and determines where transactions are processed by executing a programming code (known as chaincode) written in Go, Java or JavaScript programming languages, in the following execute-order-validate flow:

1. **Execution Phase**. A client application sends a transaction proposal to endorsing peers as specified by the relative endorsement policy, in order to invoke a chaincode function with regard to interacting with the blockchain ledger. As soon as the endorsers have successfully executed the chaincode, the endorsement is sent back to the client. A transaction is then assembled and signed with credentials obtained from a Membership Service Provider (MSP), which is a party that handles all the identities of peers and validators within each organization.

2. **Ordering Phase**. A client then sends constructed transactions to the ordering service. This is a collection of nodes known as orderers, which effectively combine multiple transactions into a single block and then broadcast the ordered transactions to all peers in the network.

3. **Validation Phase**. Lastly, each peer verifies received transactions with regard to the endorsement policy and then updates the local ledger state.

Specifically, Hyperledger Fabric demonstrates unique security mechanisms such as private data collections, which allow certain authorized participants to access only specific data [16]. To store data, peers use local state databases [14,16] such as CouchDB, MongoDB and LevelDB [15]. Participants in Hyperledger Fabric can be in the form of Docker containers. These containers [17] are a unit of software that packages programming code and dependencies in order for an application to be executed efficiently and reliably in a computing environment. Furthermore, applications are implemented in an isolated and safe surrounding, as Docker provides strong security capabilities. Several technologies can be implemented on top of Hyperledger Fabric in order to enhance its security and privacy. This includes

the Idemix cryptographic protocol suite [18], which uses ZKP to provide anonymity and unlinkability. Idemix is associated with identity certificates that each participant needs to use for performing each and every action on the distributed ledger.

3. Related Work

Electronic health records (EHR) often contain highly sensitive healthcare data, which are periodically distributed among healthcare providers, pharmacies and patients for clinical diagnosis and treatment [2]. Furthermore, critical medical information must be regularly updated and shared where proper consent is provided by the patient. Along with this we need strong availability, fast access and the appropriate encryption of these records [2,19].

There are currently several approaches regarding EHR management and how blockchain technology can be utilized to improve it [20,21]. Yüksel et al. [6] discussed the distinction of EHR storage between distributed and cloud design, and propose that a centralized model could be established by applying the relevant decentralized settings and techniques of the cloud architectures. Overall, cloud architecture refers to the structured storage and allocation of massive volumes of medical records among remote third-party service providers. Within the Cloud, healthcare organizations and individuals are able to access the data by utilizing relevant security measures based on identity management, access control, policy integration and compliance management, thus accomplishing common requirements such as scalability, availability and cost-effectiveness [1].

Sharma, Chen and Sheth [22] examined kHealth, a practical privacy-preserving cloud-based system that handles health data acquired from Internet of Things (IoT) devices. Their system aims to build personalized health predictive models by applying efficient but computationally expensive privacy guarantees, since it employs various homomorphic encryption and differential privacy techniques. Meanwhile, it should be noted that scalability optimizations undermine privacy-protection mechanisms.

Dubovitskaya et al. [8] introduced a scalable and privacy-preserving e-Health cloud system where medical data is encrypted through public-key cryptography over both local and cloud-stored databases, and efficiently distributed by using access control policies specified by the patient. A limitation of this implementation is the possible misconduct of trusted cloud server providers; these could violate anonymity by deducing private information from the user's IP address. They could also associate a pseudonym to a patient by launching an inference attack.

Casino et al. [23] addressed how blockchain technology could enhance several applications within the healthcare industry, such as the management of EHR, drug counterfeiting and user-oriented medical research.

Ming and Zhang [24] proposed an efficient privacy-preserving access control (PPAC) scheme for EHR cloud-based systems which utilizes the cuckoo filter, a novel attribute-based signcryption (ABSC) mechanism, to achieve anonymity and computational efficiency. They provided extensive privacy guarantees and comparative performance evaluation results. However, compliance with the GDPR has not been investigated.

Roehrs et al. [25] distributed personal health record (PHR) information into data blocks. From a logical point-of-view, the data storage seems centralized, but in fact it is decentralized among the participating devices. The authors noted that their proposed protocol *openPHR* is feasible, extensible, elastic and can be adopted in practice by many organizations. Their architecture is presented in detail, but the practicality of their work is being questioned. Additionally, the authors mentioned that security and privacy are still lacking in their approach. It should be noted that a PHR is controlled by the patient in contrast to an EHR, which is controlled by a healthcare institution. However, both EHRs and PHRs are electronically stored and distributed and thus may be evaluated in terms of performance and scalability metrics, privacy-preserving features and GDPR compliance.

Approaches that use a similar backbone technology to ours include [26,27]. Ichikawa et al. [26] presented solutions that utilize the permissioned blockchain Hyperledger Fabric, in order to store healthcare records collected from mobile devices. However, the authors built their model on the older

v0.5 version of Hyperledger Fabric, which does not have the private data collection feature that the v1.4 version (that PREHEALTH is built upon) has. Moreover, their model does not support the Idemix suite to create the necessary privacy guarantees. Overall, their system is able to store data in an immutable ledger but without any privacy protection for end-users. It should be noted that to incorporate the private data collection feature, an update of their system is not possible without a complete re-design of their architecture. Likewise, Liang et al. [27] utilized Hyperledger Fabric to simulate a real-world scenario with different participating entities. In their system, the represented entities include users, wearable devices, healthcare providers, insurance companies, blockchain networks and cloud databases. However, later versions of Hyperledger Fabric introduced new challenges that their work needs to address, in case of an architecture re-design and revision into a newer version that incorporates the private data collection feature. The authors conclude with helpful metrics of the feasibility of each query on their system, although important technical details are missing.

MeDShare [28] has many concepts similar to our work, however the backbone blockchain framework is not explicitly selected. Moreover, the authors focus more on discussing the fundamental building blocks of the blockchain technology, such as data blocks and smart contracts, than on a proposed solution.

The MedRec software [29] is a permissionless blockchain implementation, built on the Ethereum blockchain network, which demonstrates auditable and decentralized EHR management. It introduces both questionable transaction scalability and linkability likelihood between a transaction and an individual, thus compromising data and user anonymity. MedRec has inspired other authors to build upon it and try to mitigate its issues [30].

The MediBchain system [31] establishes a permissioned peer-to-peer approach and is designed on the Ethereum network. It uses a cloud-based server infrastructure to achieve acceptable scalability. However, even if data held by the participating parties remain in encrypted form, linkability remains an issue. Another concern is the transaction cost, which is not measurable. Another solution built on the Ethereum network is one that focuses on the pharmaceutical supply-chain: *Modum.io* [32]. In this work, Bocek et al. [32] developed an architecture where the IoT devices collect temperature data, and mobile devices get this data and connect to an HTTP server. The HTTP server acts as a blockchain node and stores the data in PostgreSQL databases, using smart contracts. Their real-world scenario looks promising and can be adapted in an EHR use case, although it is not decentralized in each stage. Where the HTTP server is hijacked by attackers or malfunctions, the collected data is vulnerable and exposed. Subsequently, its usage in an EHR scheme is not advisable, considering the sensitive nature of EHR data.

A significant and prominently distributed database technology that elaborates the blockchain technology is Blockstack [33]. Blockstack operates by default using the Gaia distributed database that is able to store its data decentralized in the users' web browsers instead of on a centralized web server, thus enhancing privacy. The Blockstack framework is currently built on the Bitcoin blockchain, but can be moved onto another platform.

Last, a decentralized privacy-preserving blockchain solution built on a permissionless blockchain network, which combines cloud storage and medical data acquired from IoT devices, is examined in [34] in the context of developing a novel hybrid EHR management method. The key advantages of the proposed solution include the utilization of lightweight symmetric and asymmetric key encryption in order to achieve effective anonymity and client authorization; however, performance efficiency and GDPR compliance have not been examined.

PREHEALTH's Features

PREHEALTH is different from previous approaches as it enables an effective and secure handling operation of highly sensitive health data over Hyperledger Fabric's permissioned blockchain network. Hyperledger Fabric is suitable for managing health records as it provides a confidential, scalable and highly flexible infrastructure solution [35], by virtue of the layers of permission and access

control that it offers. Another advantage of PREHEALTH is that its privacy-preserving features are compliant with the GDPR [36], in particular as it supports the *right to be forgotten*, by means of levels of *data anonymization*, and also by allowing stored records to have a defined *lifetime*. According to the established policy, when the defined *lifetime* expires, records can be purged from the state database, and only their hashes will remain on the immutable ledger as proof that the records existed. Similarly, participants who decide to delete their data can send a deletion transaction request to the blockchain network [37]. Likewise, a hash proving that the data existed in the ledger would remain; however, correlation with the identity of the data-owner is computationally infeasible within reasonable time limits.

Comparisons of PREHEALTH against those works in the literature whose presentations include either proof-of-concept implementations that can be utilized in real-world scenarios or detailed proposals are shown in Table 1. In this table, the underlying technology of each proposal is presented, alongside Checkmarks or Xmarks that indicate the effective privacy-preserving features, the compliance with the GDPR and whether performance and scalability evaluation results such as those presented in Section 3 have been provided.

Table 1. Comparison of related proposals.

Method	Technology	Access	Verifiability	Privacy-Preserving	GDPR	Performance/Scalability
[24]	AC Scheme	Private	Private	✓	✗	✓
[32]	Ethereum	Private	Public	✗	✗	✗
[31]	Ethereum	Private	Public	✗	✗	✓
[29]	Ethereum	Open	Public/ Private	✓	✗	✗
[33]	Bitcoin/ Agnostic	Open	Public	✓	✓	✗
[28]	Agnostic	Open	Private	✗	✗	✓
[25]	Peer-to-peer	Private	Private	✗	✗	✓
[27]	HLF	Private	Private	✗	✗	✓
[26]	HLF	Private	Private	✗	✓	✗
Our work	HLF	Private	Private	✓	✓	✓

4. Experimentation Environment

The key focus of the work is to create a system that offers anonymity and security, and acceptable scalability. To this end, it was decided that multiple Docker containers, each incorporating a core element of Hyperledger Fabric's network, would be operated in a host environment, as illustrated in Figure 2. In particular, a Debian *Stretch* host environment [38] manages a dockerized command-line interface (CLI) in order to set up a Hyperledger Fabric network consisting of peers and orderers that have installed the chaincode and the distributed Couch DB ledgers [14,15].

The overview for each distributed ledger action in Hyperledger Fabric is depicted in Figure 3. Three different organizations are deployed, each consisting of three peers and a unique MSP. Each peer maintains a copy of the public ledger, except those from the peers of *Organization 1* that manage a private data collection. This aims to demonstrate a scalable and privacy-preserving EHR management solution. As described in Section 2, the core functionality of the proposed testbed is:

- Once a peer is validated by the MSP, it sends the proposed transaction to the ordering service.
- The ordering service validates the transaction according to the associated chaincode and updates the public ledger.
- Public ledger changes are broadcast to all the peers in order to verify, accept and update their local copy of the ledger.
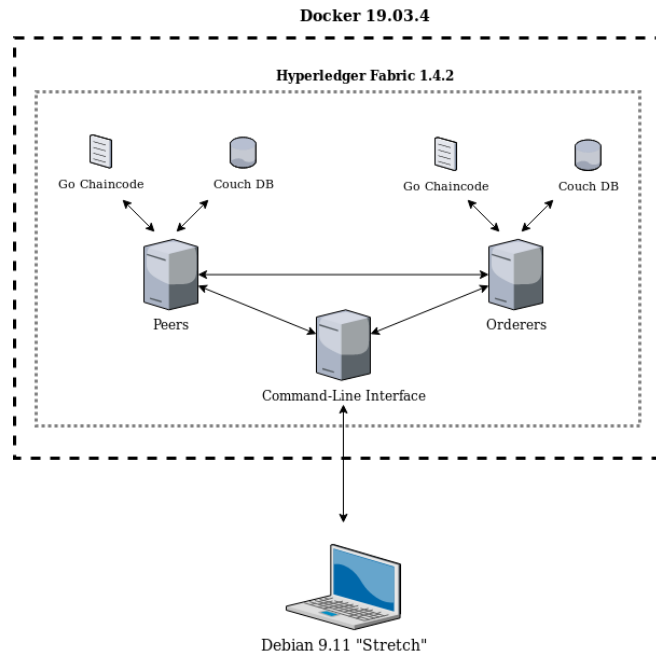
Figure 2. Proof-of-concept implementation overview.

The technical specifications of the proof-of-concept deployment are as follows: 2.4GHZ quad-core Intel Core i7 6th Generation CPU, with 8GB RAM and 256GB SSD. In particular, each data record consists of six fields, of which two represent sensitive patient data forming the private data collection and are encrypted. Specifically, the fields are distinguished in the *Name* of the patient, the *Address* of the patient, patient's *Country* of residence, patient's *Date of Birth*, a *Test* field that can include data related with each use case, and finally a generated system-specific *ID*. Moreover, the private collection data consists of the first two fields, the *Name* and the *Address* of the patient, and is accessible only by *Organization 1*, while *Organizations 2 and 3* may access only the remaining non-sensitive data. Lastly, *Read* and *Write* query times of peers are measured as part of the evaluation of the proposed solution.
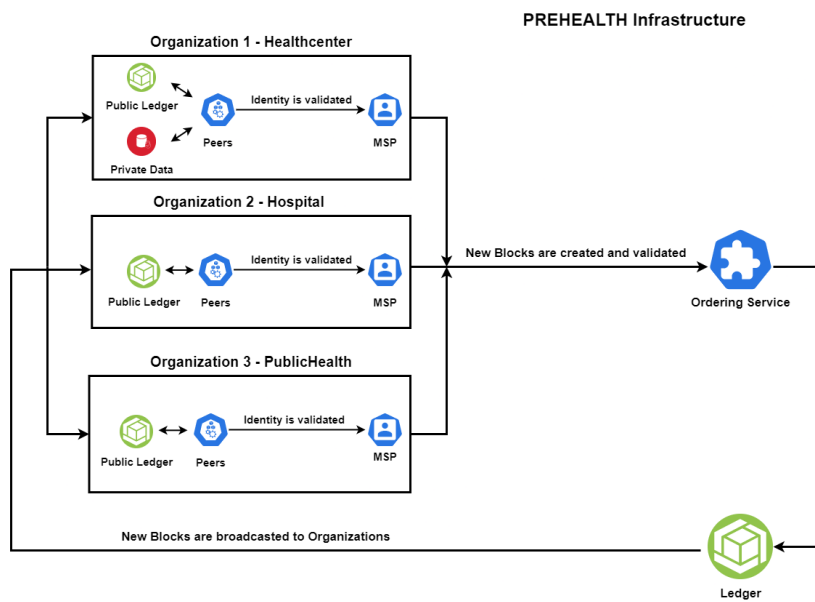

Figure 3. PREHEALTH overview.

5. Evaluation

5.1. Security Evaluation

It is common when it comes to technologies that involve identities and certificates, such as Hyperledger Fabric, that the weakest link is the human; this is inherited from traditional IT systems. In cases where the private certificates of a blockchain user are being disclosed or stolen, it is possible for a malicious user to perform arbitrary *Read* and *Write* queries to the blockchain ledger. According to the established policy, the security of the system may be compromised. In our case, this scenario is unrealistic, since no entity is completely authorized or has control over the ledger, not even the administrators. Furthermore, another risk for the blockchain systems may be the chaincode that is being executed in all the peers. Chaincode runs autonomously without external interventions, but it may contain bugs. Extensive inspection and testing are crucial for the identification of bugs in the pre-production phase [14]. Last, as the blockchain technology is still considered a new technology, while offering new and robust security mechanisms that were impossible in the past, it is still possible that further bugs and attacks may be discovered in the future. One potential threat is quantum computing, which has been developed extensively over the last few years. All the approaches relying on hashing methods and the traditional encryption mechanisms may be at risk due to the aforementioned threat. Hence, a proactive countermeasure is to use quantum-robust methodologies if they are practically feasible [39].

5.2. Anonymity Evaluation

The evaluation regarding the anonymity and unlinkability between state-of-art clients and transactions is focused on extracting relevant information during chaincode operation and blockchain user interaction. In particular, disclosed identity attributes in both common user and Idemix-based client interactions are analyzed using the client identity library [40] embedded in the chaincode. It should be noted that the aforementioned library provides an exclusive method of acquiring private information about the validators as the defined communication channel utilizes the transport layer security (TLS) cryptographic protocol, thus efficiently encrypting relevant information, while blockchain transactions are only signed by a cryptographic public-key, which does not reveal confidential data.

Furthermore, in the work of [15], the authors noted that Hyperledger Fabric provides a broad configuration space; therefore, particular parameters should be adjusted for anonymity and blockchain network performance measurements to be successfully determined and assessed. These components are divided into static and dynamic variables, and are analyzed in the current section as follows.

Static variables:

1. **Number of organizations, peers and orderers**—Considering complexity and technical limitations, three organizations were established, consisting of three peers each acting as validators and an ordering service of three orderers.
2. **Endorsement policy**—Each transaction proposal requires at least one peer of any corresponding organization to sign the transaction request, thus eliminating unnecessary communication between endorsing parties.

Dynamic Variables:

1. **Membership service provider**—Idemix technology requires a distinct authentication provider in contrast to the prevailing X.509 public key certificate mechanism [41].
2. **Registration scheme for users**—Unique CLI commands are required in the context of Idemix user registration. Specifically, a suitable CLI docker container was developed, as presented in Figure 2, in which relevant Idemix parameters were handled as command-line arguments in order for a user to register and interact with the blockchain network.

5.3. Performance Evaluation

We calculated query times for PREHEALTH, as well as for a PostgresSQL database simulating a common EHR management solution, and for the MedRec and Blockstack experimentation prototypes for various volumes of EHRs (10, 100, 1000, 10,000, 100,000 and 1,000,000). The results are presented in Table 2 and in Figure 4. PREHEALTH demonstrates efficient *Read* and *Write* query time measurements while both blockchain solutions, namely Medrec and Blockstack, but also a traditional PostgreSQL database, impose higher overhead per query over time. Notably, PREHEALTH exhibits much better results when the number of records is large. As can be seen from the conducted tests, the transaction time for each *Read* query using a PostgreSQL database is rising linearly and is expected to overcome PREHEALTH at approximately 1,200,000 records. It is clear, then, that in a real-world scenario where there are millions of stored records, PREHEALTH can handle each query effectively and efficiently.

Furthermore, it should be noted that the MedRec prototype has been developed in an experimental environment by utilizing the Proof of Authority consensus algorithm, in order for local nodes to be successfully deployed. This is in contrast to the operation of a time-consuming Proof of Work consensus model with regard to the commercial MedRec application [29], thus leading to increased query time measurements.

Table 2. Query time measurements in milliseconds (ms) per number of entries.

| Number of Records:		10	100	1000	10,000	100,000	1,000,000
PREHEALTH	Read Data Time	183 ms	183 ms	183 ms	183 ms	183 ms	183 ms
	Write Data Time	58 ms	58 ms	58 ms	58 ms	58 ms	58 ms
PostgresSQL Database	Read Data Time	1.73 ms	1.79 ms	2.38 ms	8.76 ms	43.52 ms	136.19 ms
	Write Data Time	4.32 ms	4.48 ms	4.47 ms	4.37 ms	4.39 ms	4.45 ms
MedRec—Azaria et al. [29]	Read Data Time	177 ms	186 ms	194 ms	199 ms	205 ms	210 ms
	Write Data Time	81.5 ms	86.9 ms	79.6 ms	71.6 ms	63.2 ms	79.6 ms
Blockstack—Ali et al. [33]	Read Data Time	360 ms	360 ms	360 ms	360 ms	360 ms	360 ms
	Write Data Time	530 ms	530 ms	530 ms	530 ms	530 ms	530 ms
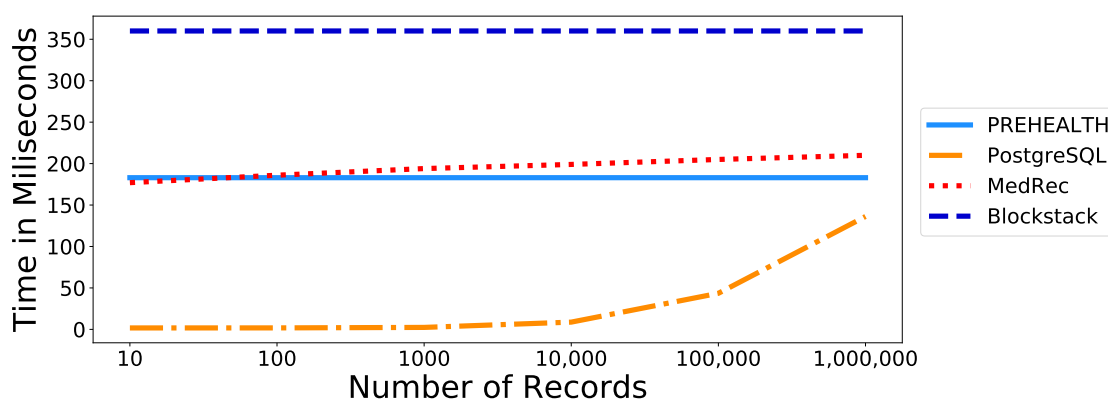
Figure 4. Read records transactions overhead.

Furthermore, benchmark tests were conducted for the *CPU* and *Memory* performance of our proposed PREHEALTH solution. The CPU usage of PREHEALTH peers during the workflow for *Read* and *Write* queries over various volumes of EHRs (1000, 10,000, 100,000) is presented in Table 3 and Figure 5. As depicted in Figure 3, PREHEALTH introduced three organizations, namely *Healthcenter*, *Hospital* and *PublicHealth*. Each organization comprises three blockchain peers, respectively *Peer 0*, *Peer 1* and *Peer 2*. The average CPU usage of the blockchain peers during *Read* queries is reasonable and less than 30%, whilst the CPU performance during *Write* queries is even lower, with a maximum

of 15.4%. The CPU performance of the blockchain peers is regularly stable with swift fluctuations that are presumably caused by hardware limitations. The average memory overhead of each blockchain peer during our tests over various volumes of EHRs was insignificant (less than 2%); hence, it is not illustrated in a plot.

Table 3. Average CPU (%) performance of each blockchain peer per number of electronic health records.

| PREHEALTH Organizations | PREHEALTH Peers | | Number of Records | | |
			1000	10,000	100,000
Healthcenter	Peer 0	Read Queries	7.6%	28.7%	29%
		Write Queries	6.7%	10.3%	15.4%
	Peer 1	Read Queries	5.1%	21.8%	21.7%
		Write Queries	4.9%	6.7%	4.2%
	Peer 2	Read Queries	4.9%	23.3%	22.2%
		Write Queries	5.4%	6.4%	4.3%
Hospital	Peer 0	Read Queries	8.3%	29.4%	32.3%
		Write Queries	9.3%	11.2%	13.9%
	Peer 1	Read Queries	5.1%	22.7%	23.2%
		Write Queries	5.4%	6.4%	4.3%
	Peer 2	Read Queries	5.4%	20.7%	18.7%
		Write Queries	4.9%	6.6%	4.2%
PublicHealth	Peer 0	Read Queries	7.6%	30.5%	30.3%
		Write Queries	11.4%	12.8%	8.2%
	Peer 1	Read Queries	4.8%	22%	20.1%
		Write Queries	5.3%	6.8%	4%
	Peer 2	Read Queries	5.1%	23.4%	22%
		Write Queries	4.7%	6.6%	4%
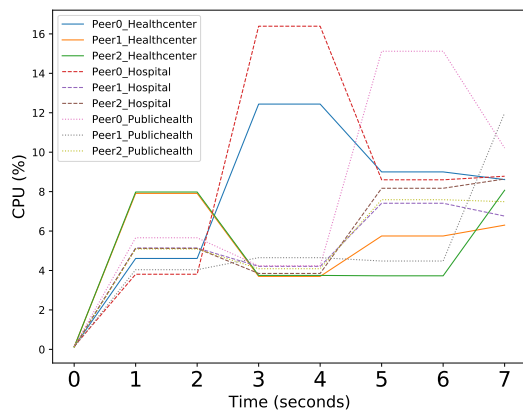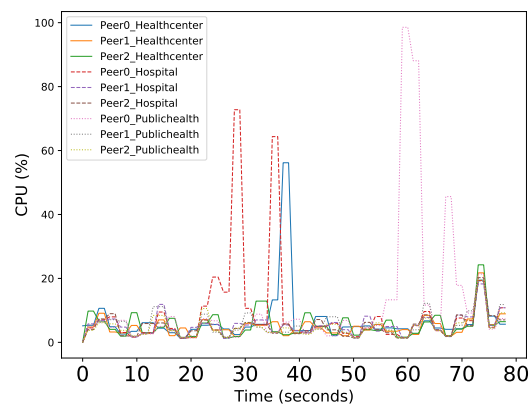
Figure 5. *Cont.*
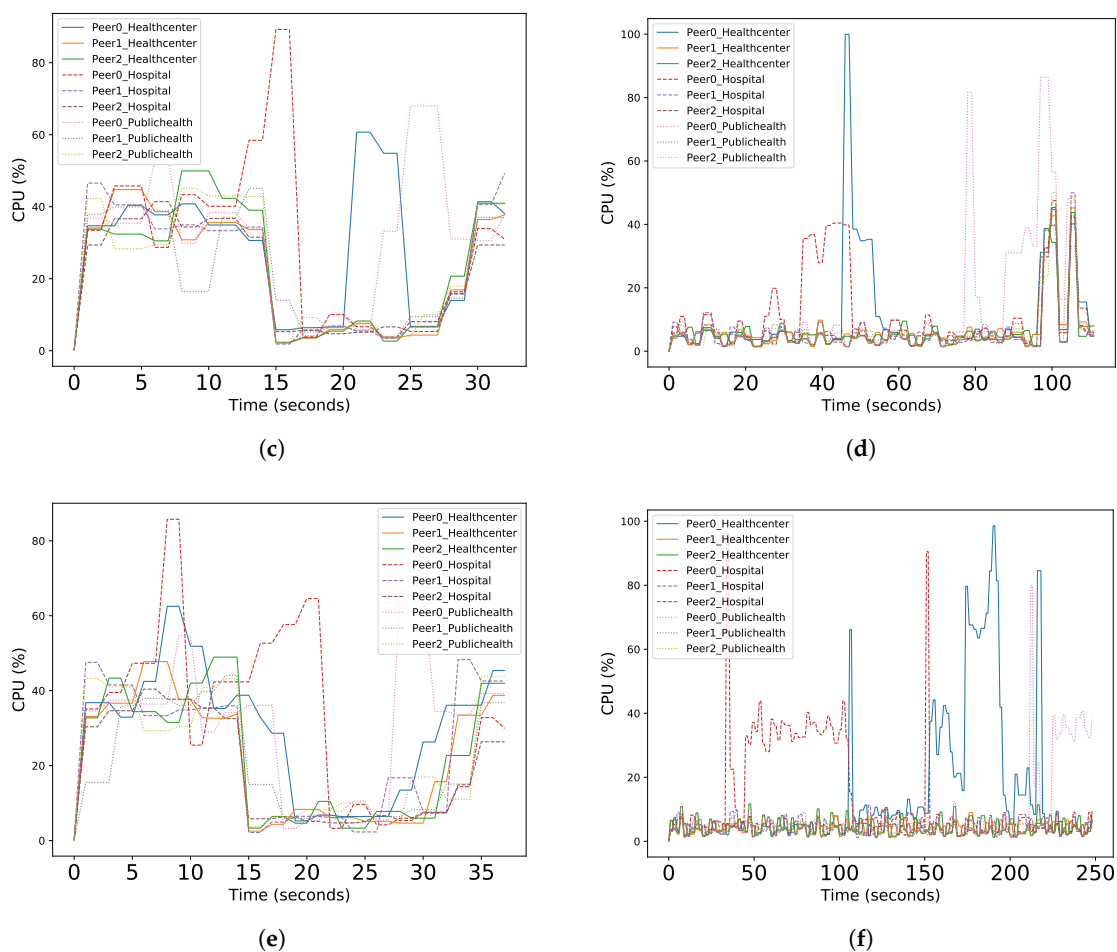
Figure 5. (**a**) Read queries workflow on 1000 Records. (**b**) Write queries workflow on 1000 Records. (**c**) Read queries workflow on 10,000 Records. (**d**) Write queries workflow on 10,000 Records. (**e**) Read queries workflow on 100,000 Records. (**f**) Write queries workflow on 100,000 Records. CPU Usage (%) of blockchain peers during workflow.

6. Discussion

An empirical comparison of PREHEALTH to the state-of-art procedures of the Hyperledger Fabric framework, which were utilized in order to achieve explicit anonymity and privacy of transactions, was conducted in the previous section. Additionally, experimental metrics in terms of blockchain performance, such as blockchain scalability, immutability, auditability and cost-effectiveness, were assessed in association with selected environmental settings and parameters. Finally, tests were reviewed in terms of deploying an experimental proof-of-concept to a case-study environment.

Calculation and examination of measurable parameters and performance metrics with regard to the experimental implementation were also considered. Apart from state-of-art Idemix technology utilization, certain blockchain features such as scalability, auditability, immutability and cost-effectiveness were successfully validated. In particular, our findings confirm that the proposed EHR management approach provides the capability for medical record transaction history to be both immutable and properly audited, even at a greater scale, without significant performance overhead.

Effective encryption mechanisms were applied to successfully encrypt information; as a result, a malicious third party cannot access sensitive data. Furthermore, a blockchain ledger is invulnerable to specific command injection threats such as SQL injection, which targets a specific format of an information database [42]. Moreover, the MSP directory containing appropriate cryptographic material was mounted in each relative blockchain peer, thus rendering infeasible the exploitation of crafted

cryptographic components by an attacker attempting to impersonate a blockchain peer and thus compromise the assembled network. Lastly, the implementation was associated with a private isolated blockchain topology, invulnerable to external security threats, where identified blockchain peers and orderers would directly reveal their identity in the case of ill-behavior. In the case of a permissioned private blockchain consortium that would use PREHEALTH, all validator identities are disclosed. Therefore, upon attempting to tamper with EHRs or blockchain transactions themselves, the actor (i.e., validator) would reveal their identity.

7. Conclusions and Future Work

The management of EHRs is and will continue to be crucial for years to come. EHRs consist of sensitive health data of subjects whose privacy must be protected. We proposed PREHEALTH, an approach that can be used to store EHRs while ensuring the patients' privacy. We provided an empirical comparison of our proof-of-concept implementation against existing alternative proposals in the literature, and found that PREHEALTH provides more privacy guarantees and at the same time imposes less query overhead when the volume of the stored data is high. Indeed, PREHEALTH achieves a read query time of 183 ms in any number of EHRs, when the time for a read query in the other compared works is higher or rising exponentially. Specifically, a read query on the MedRec platform takes 210 ms in 1,000,000 EHRs, in Blockstack it takes 360 ms and a PostgreSQL database performs a read query in 136.19 ms. In addition, we have proposed and used measurable parameters for evaluating the scheme's performance, scalability, auditability and immutability.

A possible extension of this work could be focused on operating an advanced version of Idemix credentials, in a future release. Once credential revocation and incorporation of custom identity attributes are enabled, technical limitations of the Idemix technology would be resolved and a practical exploratory activity could be undertaken in order for a production privacy-preserving EHR handling solution to be developed. In addition, another approach with regard to performance benchmarking could be a relative blockchain solution in a Kubernetes cluster, while a certain number of configurable parameters could be examined in association with their effects on blockchain scalability. Specifically, a high-performance cloud environment could be utilized for the implementation of a complex blockchain network with the incorporation of numerous blockchain peers, orderers, endorsement policies and parallel blockchain interactions, thus efficiently evaluating metrics along with examining the scalability as a whole.

Author Contributions: All authors contributed to the conceptualization and methodology of the manuscript; C.S. performed the data preparation; P.P. and N.P. contributed to writing; S.K. and W.J.B. reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abbas, A.; Khan, S.U. A review on the state-of-the-art privacy-preserving approaches in the e-health clouds. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 1431–1441. [CrossRef] [PubMed]
2. Dubovitskaya, A.; Xu, Z.; Ryu, S.; Schumacher, M.; Wang, F. Secure and trustable electronic medical records sharing using blockchain. *AMIA Annu. Symp. Proc.* **2017**, *2017*, 650–659. [PubMed]
3. Coventry, L.; Branley, D. Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. *Maturitas* **2018**, *113*, 48–52. [CrossRef] [PubMed]
4. Smart, W. *Lessons Learned Review of the WannaCry Ransomware Cyber Attack*; Department of Health and Social Care: London, UK, 2018; Volume 1, pp. 10–1038.
5. Alvarez, M. *Security Trends in the Healthcare Industry*; Somers IBM: Somers, NY, USA, 2017; pp. 2–18. Available online: https://www.ibm.com/downloads/cas/PLWZ76MM (accessed on 17 November 2020).

6. Yüksel, B.; Küpçü, A.; Özkasap, Ö. Research issues for privacy and security of electronic health services. *Future Gener. Comput. Syst.* **2017**, *68*, 1–13. [CrossRef]

7. Renuka, K.; Kumar, S.; Kumari, S.; Chen, C.M. Cryptanalysis and improvement of a privacy-preserving three-factor authentication protocol for wireless sensor networks. *Sensors* **2019**, *19*, 4625. [CrossRef] [PubMed]

8. Dubovitskaya, A.; Urovi, V.; Vasirani, M.; Aberer, K.; Schumacher, M.I. A cloud-based ehealth architecture for privacy preserving data integration. In Proceedings of the IFIP International Information Security and Privacy Conference, Hamburg, Germany, 26–28 May 2015; pp. 585–598.

9. Angelis, S.D.; Aniello, L.; Baldoni, R.; Lombardi, F.; Margheri, A.; Sassone, V. PBFT vs proof-of-authority: Applying the CAP theorem to permissioned blockchain. In Proceedings of the Italian Conference on Cyber Security, Milan, Italy, 6 February 2018.

10. Kuo, T.T.; Kim, H.E.; Ohno-Machado, L. Blockchain distributed ledger technologies for biomedical and health care applications. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 1211–1220. [CrossRef] [PubMed]

11. Podda, A.S.; Pompianu, L. An overview of blockchain-based systems and smart contracts for digital coupons. In Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, Seoul, Korea, 24 June–16 July 2020; pp. 770–778.

12. Longo, R.; Podda, A.S.; Saia, R. Analysis of a Consensus Protocol for Extending Consistent Subchains on the Bitcoin Blockchain. *Computation* **2020**, *8*, 67. [CrossRef]

13. Cachin, C. Architecture of the hyperledger blockchain fabric. In Proceedings of the Workshop on Distributed Cryptocurrencies and Consensus Ledgers, Chicago, IL, USA, 25 July 2016; Volume 310, p. 4.

14. Androulaki, E.; Barger, A.; Bortnikov, V.; Cachin, C.; Christidis, K.; De Caro, A.; Enyeart, D.; Ferris, C.; Laventman, G.; Manevich, Y.; et al. Hyperledger fabric: A distributed operating system for permissioned blockchains. In Proceedings of the Thirteenth EuroSys Conference, Porto, Portugal, 23–26 April 2018; pp. 1–15.

15. Thakkar, P.; Nathan, S.; Viswanathan, B. Performance benchmarking and optimizing hyperledger fabric blockchain platform. In Proceedings of the 2018 IEEE 26th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Milwaukee, WI, USA, 25–28 September 2018; pp. 264–276.

16. Papadopoulos, P.; Pitropakis, N.; Buchanan, W.J.; Lo, O.; Katsikas, S. Privacy-Preserving Passive DNS. *Computers* **2020**, *9*, 64. [CrossRef]

17. Boettiger, C. An introduction to Docker for reproducible research. *ACM SIGOPS Oper. Syst. Rev.* **2015**, *49*, 71–79. [CrossRef]

18. Androulaki, E.; De Caro, A.; Neugschwandtner, M.; Sorniotti, A. Endorsement in Hyperledger Fabric. In Proceedings of the 2019 IEEE International Conference on Blockchain (Blockchain), Atlanta, GA, USA, 14–17 July 2019; pp. 510–519.

19. Abramson, W.; Hall, A.J.; Papadopoulos, P.; Pitropakis, N.; Buchanan, W.J. A Distributed Trust Framework for Privacy-Preserving Machine Learning. *arXiv* **2020**, arXiv:2006.02456.

20. Hölbl, M.; Kompara, M.; Kamišalić, A.; Nemec Zlatolas, L. A systematic review of the use of blockchain in healthcare. *Symmetry* **2018**, *10*, 470. [CrossRef]

21. Mayer, A.H.; da Costa, C.A.; da Rosa Righi, R. Electronic health records in a blockchain: A systematic review. *Health Inform. J.* **2020**, *26*, 1273–1288. [CrossRef] [PubMed]

22. Sharma, S.; Chen, K.; Sheth, A. Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems. *IEEE Internet Comput.* **2018**, *22*, 42–51. [CrossRef]

23. Casino, F.; Dasaklis, T.K.; Patsakis, C. A systematic literature review of blockchain-based applications: current status, classification and open issues. *Telemat. Inform.* **2019**, *36*, 55–81. [CrossRef]

24. Ming, Y.; Zhang, T. Efficient privacy-preserving access control scheme in electronic health records system. *Sensors* **2018**, *18*, 3520. [CrossRef] [PubMed]

25. Roehrs, A.; da Costa, C.A.; da Rosa Righi, R. OmniPHR: A distributed architecture model to integrate personal health records. *J. Biomed. Inform.* **2017**, *71*, 70–81. [CrossRef] [PubMed]

26. Ichikawa, D.; Kashiyama, M.; Ueno, T. Tamper-resistant mobile health using blockchain technology. *JMIR MHealth UHealth* **2017**, *5*, e111. [CrossRef] [PubMed]

27. Liang, X.; Zhao, J.; Shetty, S.; Liu, J.; Li, D. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. In Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 8–13 October 2017; pp. 1–5.

28. Xia, Q.; Sifah, E.B.; Asamoah, K.O.; Gao, J.; Du, X.; Guizani, M. MeDShare: Trust-less medical data sharing among cloud service providers via blockchain. *IEEE Access* **2017**, *5*, 14757–14767. [CrossRef]

29. Azaria, A.; Ekblaw, A.; Vieira, T.; Lippman, A. Medrec: Using blockchain for medical data access and permission management. In Proceedings of the 2016 2nd International Conference on Open and Big Data (OBD), Vienna, Austria, 22–24 August 2016; pp. 25–30.

30. Yang, H.; Yang, B. A blockchain-based approach to the secure sharing of healthcare data. In Proceedings of the Norwegian Information Security Conference 2017, Oslo, Norway, 11–15 September 2017.

31. Al Omar, A.; Rahman, M.S.; Basu, A.; Kiyomoto, S. Medibchain: A blockchain based privacy preserving platform for healthcare data. In Proceedings of the International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage, Guangzhou, China, 12–15 December 2017; pp. 534–543.

32. Bocek, T.; Rodrigues, B.B.; Strasser, T.; Stiller, B. Blockchains everywhere-a use-case of blockchains in the pharma supply-chain. In Proceedings of the 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Lisbon, Portugal, 8–12 May 2017; pp. 772–777.

33. Ali, M.; Shea, R.; Nelson, J.; Freedman, M.J. Blockstack: A New Decentralized Internet. Whitepaper. May 2017. Available online: https://pdos.csail.mit.edu/6.824/papers/blockstack-2017.pdf (accessed on 17 November 2020).

34. Dwivedi, A.D.; Srivastava, G.; Dhar, S.; Singh, R. A decentralized privacy-preserving healthcare blockchain for IoT. *Sensors* **2019**, *19*, 326. [CrossRef] [PubMed]

35. Albeyatti, A. White Paper: Medicalchain. MedicalChain Self-Publication, 2018. Available online: https://medicalchain.com/en/whitepaper/ (accessed on 17 November 2020).

36. Voigt, P.; Von dem Bussche, A. The eu general data protection regulation (gdpr). In *A Practical Guide*, 1st ed.; Springer International Publishing: Cham, Switzerland, 2017.

37. Davari, M.; Bertino, E. Access control model extensions to support data privacy protection based on GDPR. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 4017–4024.

38. Krafft, M.F. *A Delphi Study of the Influences on Innovation Adoption and Process Evolution in a Large Open Source Project: The Case of Debian*; University of Limerick, Department of Computer Science & Information Systems: Limerick, Ireland, 2010. Available online: http://193.1.102.107/handle/10344/427 (accessed on 17 November 2020).

39. English, E.; Kim, A.D.; Nonaka, M. Advancing Blockchain Cybersecurity: Technical and Policy Considerations for the Financial Services Industry. 2018. Available online: https://www.microsoft.com/en-us/cybersecurity/content-hub/advancing-blockchain-cybersecurity (accessed on 17 November 2020).

40. Hyperledger Fabric. Client Identity Chaincode Library, 2017. Available online: https://github.com/hyperledger/fabric/blob/release-1.1/core/chaincode/lib/cid/README.md (accessed on 17 November 2020).

41. Chokhani, S.; Ford, W.; Sabett, R.; Merrill, C.; Wu, S. *RFC 2527: Internet X.509 Public Key Infrastructure Certificate Policy and Certification Practices Framework*; Internet Engineering Task Force (IETF), RFC: Fremont, CA, USA, 1999. Available online: https://tools.ietf.org/html/rfc2527 (accessed on 17 November 2020).

42. Halfond, W.G.; Viegas, J.; Orso, A. A classification of SQL-injection attacks and countermeasures. In Proceedings of the IEEE International Symposium on Secure Software Engineering, Washington, DC, USA, March 2006; Volume 1, pp. 13–15.