1 2 9 0

UNIVERSIDADE D
COIMBRA

Criptography

Managing Digital Identity in Education

Francisco Catarino Mendes - 2019222823
Department of Informatics Engineering
University of Coimbra

Introduction

Nowadays, blockchain implementations are becoming more and more present in the technology industry, and some may say that everything that revolves around blockchain can be characterized as the future. Blockchain is a shared, immutable ledger that facilitates the process of recording transactions and tracking assets in a business network [1].

In this paper, the goal is to understand how digital identity in education can be managed using a blockchain tool, giving a description of this tool and why it was chosen, specifying the selected use case of digital identity in education, describing how its implementation occurs, and finally analyzing the pros and cons of the in regards to the success of the implementation.

1 Used Blockchain Tool - Hyperledger Fabric

As for the blockchain tool, Hyperledger Fabric was selected. Hyperledger Fabric, an open-source project from the Linux Foundation, is the modular blockchain framework and de facto standard for enterprise blockchain platforms. It has advanced privacy controls so only the data you want shared gets shared among the known network participants [2].

Due to the fact that it is a permission blockchain network, Hyperledger can mainly solve confidentiality problems that organizations usually face by offering the facility to create a blockchain application maintaining the privacy of the information of the company [3]. It allows for identity and user management through organization networks, and lists of access controls add more granular permissions to enhance privacy and reduce the possibility of breaches [4]. These were the main reasons behind the choice of Hyperledger, since when managing digital identity, confidentiality and integrity are key.

2 Selected Use Case - Digital Identity in Education

Companies, like educational institutions, often collect sensitive information about their users and store it alongside less sensitive routine business data. This creates new business risks with the rise of user privacy-centric regulations such as GDPR and the shifting industry focus to corporate IT responsibility. Only after receiving large fines or developing stronger IT capabilities will many enterprises pursue expensive and risky projects to achieve the right balance between data security and business needs [5]. In regard to the education world, student data is extremely sensitive data that can cause major problems if not handled with the right care. This makes the area a perfect candidate to adopt decentralized digital identity mechanisms, with the help of blockchain.

2.1 Problems with Current Digital Identity Management Systems

A digital identity is the total information about an individual or organization that exists online [6]. The current digital identity management systems are centralized and federated, with numerous single points of failure, containing some weaknesses that lead to the following concerns:

(a) **Risk of Data Breaches** - because so much user data is kept in one place, hackers could potentially access a large amount of information, leading to data breaches;

(b) **Managing many login accounts** - every time someone wants to sign up to a different system, new credentials are generated, increasing security risk;

(c) **Identity Theft** - can be a consequence of data breaches, the weak link between digital and offline identities makes it relatively easy to create fake identities;

(d) **Lack of Data Control and Ownership**;

(e) **Inaccessibility to Official Identity**.

2.2 Digital Identity in Education with Hyperledger Fabric

Implementing digital identity management on Hyperledger Fabric for the education sector involves creating a secure and transparent system for verifying and maintaining student records, as well as managing teachers' data.

Students are attributed with a unique digital identity within the blockchain (DID). DIDs are decentralized identifiers secured by a private key. With the DIDs, they can access stored encrypted student information, including personal details, academic records, and certifications, as well as transactions related to student enrollment, course completion, certification issuance, and teacher assignments. Students can also hold their credential that confirms that they are students on their phone without even needing to reveal their personal details, in case they need to prove their identity in order to access some product discounts.

Teachers are also attributed with a unique digital identity within the blockchain (DID), from which they can access stored encrypted teacher information, including personal details, professional qualifications, and courses taught.

Finally, some smart contracts are also put in place. Smart contracts are simply programs stored on a blockchain that run when predetermined conditions are met. They typically are used to automate the execution of an agreement so that all participants can be immediately certain of the outcome, without any intermediary's involvement or time loss [7]. With that said, here are some of the smart contracts implemented:

(a) **Enrollment Contract** - a smart contract that manages the process of enrolling students in courses and programs;

(b) **Certification Contract** - smart contracts responsible for verifying and issuing academic certificates;

(c) **Teacher Assignment Contract** - smart contracts handling the assignment of teachers to specific courses or classes.

3 Use Case Implementation

3.1 Setup

Here, it is established a Hyperledger Fabric network with nodes representing educational institutions, students, teachers, and necessary administrative bodies. Setting up a Hyperledger Fabric network involves several steps, including defining the network configuration, creating the necessary cryptographic materials, and launching the network components.

An example could be:

(a) **Defining the Network Configuration** - by creating a 'network-config.yaml' file to define the network configuration;

(b) **Generating Cryptographic Materials** - by using the Hyperledger Fabric CA tool to generate cryptographic materials (certificates, keys);

(c) **Launching Network Components** - by starting the Hyperledger Fabric network components using Docker Compose ('docker-compose.yaml' file);

(d) **Running the network** (docker-compose up -d).

3.2 Smart Contracts (Chaincode)

3.2.1 Encryption Standards

As one can imagine, data will need encryption protection. In this case, Elliptic Curve Cryptography (ECC) was chosen due to its efficiency and suitability for blockchain environments, and the fact that it offers strong security with shorter key lengths compared to traditional algorithms like RSA. Then, using Hyperledger built-in tools for key generation, a pair of ECC keys is generated: a public key for encryption, which is integrated into the chaincode logic, and a private key for decryption, which should be securely stored and managed.

As for data at rest, Hyperledger Fabric, by default, secures data using cryptographic primitives. Regarding data in transit, Hyperledger uses Transport Layer Security (TLS) to secure communication between nodes.

Going further into the matter of key management, here are some of the policies implemented:

(a) Private keys are stored securely in the MSP (Membership Service Provider) directory associated with each entity;

(b) Functions are defined within the chaincode to trigger key rotation at predefined intervals or based on certain events, as well as to revoke keys associated with entities that are no longer part of the network or in case of a security incident;

(c) Integration with a Hardware Security Module, as they provide additional security by offloading key management operations to dedicated hardware;

(d) Utilization of Hyperledger Fabric's secure communication channels, which are protected by Transport Layer Security (TLS), to securely transfer public keys during enrollment and other operations.

3.2.2 Access Control

All functions of the smart contracts must have access control policies. This process goes as follows:

(a) Defining Roles and Permissions of the entities within the network, using Role-Based Access Control (RBAC);

(b) Creating Access Control Lists that specify which roles have access to read or modify specific types of data;

(c) Implementing access control logic at the beginning of each function, by checking the identity of the entity invoking the transaction;

Hyperledger makes this possible due to its capability to define fine-grained permissions based on attributes or roles.

3.2.3 Data Integrity

Moreover, data integrity should be ensured at all times. For that, a suitable cryptographic hash function must be chosen, which in this case is going to be SHA-256, commonly used for its strength and efficiency. Various hashing procedures are described below:

(a) Before committing data to the ledger, the hash of the data is calculated using SHA-256;

(b) The hash value is then appended to the data being stored;

(c) When reading data from the blockchain, recalculate the hash of the stored data and compare it with the stored hash value to see if the data has been tampered with;

(d) When creating or updating transactions, hash values are included in the transaction payload;

(e) Calculating the hash of critical sections of the chaincode logic, storing and then using them during deployment or updates to verify that the chaincode has not been tampered with.

3.3 Consensus Mechanism

A consensus mechanism is a program used in blockchain systems to achieve distributed agreement about the ledger's state. Generally, it is implemented in a network with many processes and users. Organizations benefit from their use because the consensus mechanism replaces much slower human verifiers and auditing [8].

For this project, Raft is the chosen consensus mechanism, Hyperledger Fabric's default consensus mechanism. Even though an argument could be made to implement Practical Byzantine Fault Tolerance (PBFT) instead, the fact that the participants in the network are generally trusted, Raft is generally considered easier to deploy and manage, and the size of the network is constantly growing makes Raft the best consensus mechanism for this use case.

Implementation of Raft can go as follows:

(a) Firstly, an update of the network configuration has to be made;

(b) Then, the orderer genesis block must be generated using the updated configuration;

(c) Network must be launched;

(d) Join channels and update chaincode.

Furthermore, transaction finality, which means that the transaction is highly unlikely to be reversed due to a crash or failure in the network, must be addressed when using Raft:

(a) Raft works by electing a leader among a set of nodes. The leader is responsible for receiving and ordering transactions into blocks. Once a block is proposed by the leader and accepted by a majority of nodes, it is considered committed;

(b) When a client proposes a transaction, it is first sent to an endorsing peer for endorsement. After receiving endorsements, the client sends the endorsed transaction to the Raft orderer nodes:

1 - The leader of the Raft network orders the transactions into a block;

2 - Once the leader proposes a block, it disseminates it to the other nodes in the Raft cluster.

(c) When more than half of the Raft nodes acknowledge the reception of a block, it is considered committed and added to the blockchain ledger. The leader sends a commit message to the other nodes once a block is accepted. The nodes then apply the block to their ledgers.

3.4 Integration with External Systems

Integration with external systems is fundamental so that the blockchain becomes functional and serves its purpose:

(a) **Authentication Systems** - integration with external authentication systems for both students and teachers. This may involve Single Sign-On (SSO) solutions or identity providers to enhance the security and usability of the digital identity management system. Smart contracts for authentication can also be developed to interact with the external authentication system;

(b) **Credential Verification Services** - establishing connections with external credential verification services to ensure the validity of academic records and professional qualifications. This could involve real-time verification or periodic checks against authoritative databases;

(c) **Data Validation Flow** - defining a secure and efficient flow for data validation between the Hyperledger Fabric network and external systems;

(d) **Smart Contract Interfaces** - developing standardized smart contract interfaces that facilitate seamless integration with external systems. These interfaces should define how external systems can interact with the blockchain for identity verification and data validation.

4 Pros and Cons of Hyperledger Fabric

4.1 Pros

1. **Achieving Efficiency** - Efficiency can be improved by relying on the blockchain for identity management, user authentication, and access authorization. Around 30% and 40% of customer service calls are in some way related to authentication problems. Having one identity that provides access to all the applications and data needed minimizes uncertainty and frustration [4].
2. **Interoperable Systems** - Instead of several different sets of passwords, being able to receive authorization and access using a single identity eliminates the silos and makes network access more versatile [4].
3. **Modular Architecture** - Hyperledger Fabric's modular architecture allows for flexibility and scalability. This is advantageous for adapting the blockchain network to changing requirements in the education sector.
4. **Identity Management Capabilities** - Hyperledger includes built-in identity management features, supporting the integration of different authentication mechanisms. This is essential for securely managing digital identities in an educational context.
5. **Enhanced Privacy** - Networks and private transactions are provided to tackle data protection and privacy problems associated with third-party access [4].
6. **Comprehensive Documentation** - Hyperledger Fabric has extensive documentation and an active community. This provides valuable resources for developers and administrators implementing and maintaining digital identity solutions in education.

4.2 Cons

1. **Learning Curve** - Implementing and managing this tool requires a learning curve, especially for users new to blockchain technology. Training and expertise are essential for successful deployment and maintenance.
2. **Resource Intensive** - Hyperledger Fabric can be resource-intensive, especially in terms of computation and storage. This may pose challenges for organizations with limited resources or infrastructure.

3. **Complexity of Configuration** - Configuring Hyperledger Fabric, especially for specific use cases, can be complex. Customizing the network requires careful consideration of various parameters and settings.

4. **Regulatory Considerations** - The regulatory landscape for blockchain in education is evolving. Institutions need to navigate and comply with regulatory requirements, which may vary across jurisdictions.

4.3 Analysis

All in all, Hyperledger Fabric offers a robust framework for managing digital identity in education, with an emphasis on privacy, permissions, and flexibility. However, organizations should carefully weigh the pros and cons based on their specific needs, resources, and the regulatory environment in which they operate. Successful implementation requires thorough planning, expertise, and a clear understanding of the unique challenges in the education sector, but in the end, this implementation can bring a lot of positive consequences, and with technology always advancing and people's knowledge constantly improving, this blockchain tool seems to have a bright future not only in the area of education but generally speaking.

Conclusion

This academic paper provided a description and design of how an educational institution network would function when managing its digital identities with a blockchain tool, which in this case was Hyperledger Fabric. Moreover, a summary of the tool was given, as well as a comparison between the advantages and disadvantages of using Hyperledger for this process.

Even though there is hope that this document was useful and enlightening for those who want to follow and dive deep into the overall matter of blockchain, regarding future work, the potential is considerable, since as of this phase no code was actually developed in order to give life to the overall idea. Besides, more detailed work could be done about the connection and interactions of the blockchain tool with all the external systems it associates with, as well as regarding the architecture of the whole project.

References

[1] IBM - "What is Blockchain Technology?", available at https://www.ibm.com/topics/blockchain.

[2] IBM - "What is Hyperledger Fabric?", available at https://www.ibm.com/topics/hyperledger.

[3] GeeksforGeeks - "Blockchain - Hyperledger vs Ethereum", available at https://www.geeksforgeeks.org/blockchain-hyperledger-vs-ethereum/. March 2, 2023.

[4] Oodles Blockchain - "Digital Identity Management with Hyperledger Projects", available at https://blockchain.oodles.io/blog/digital-identity-management-hyperledger/. September 14, 2020.

[5] Consensys - "Blockchain in Digital Identity", available at https://consensys.net/blockchain-use-cases/digital-identity/.

[6] Dock - ”Blockchain Identity Management: Complete Guide 2023”, available at https://www.dock.io/post/blockchain-identity-management#problems-with-current-digital-identity-management-systems. October 23, 2023.

[7] IBM - ”What are Smart Contracts on Blockchain?”, available at https://www.ibm.com/topics/smart-contracts.

[8] Frankenfield, Jake. ”What Are Consensus Mechanisms in Blockchain and Cryptocurrency?”, available at https://www.investopedia.com/terms/c/consensus-mechanism-cryptocurrency.asp. February 17, 2023.