香 港 大 學

THE UNIVERSITY OF HONG KONG

Decentralized Identity Platform by Blockchain

CAES9542 - Progress Report 1

Ankit Tibrewal (3035345784)

Shivansh Mittal (3035347330)

Supervisor: Dr. John Yuen

October 20, 2019

Abstract

In recent times, most web services depend upon centralized databases owned by big corporations to carry out user authentication and verification processes. As a result, the participating users' personal information and private credentials are not under their control. Not only does this data become susceptible to being traded among businesses, it also becomes highly vulnerable to hacks and breaches, since the hackers only have to break through a single point of failure. Thus, to solve these privacy related issues, this report introduces a solution based on the concept of decentralized identities (DIDs). This project aims to create an improved solution to existing digital identity validation mechanisms and seize control of users' sensitive information from centralized parties and hand it over to their rightful owners. Using the Ethereum blockchain, the solution enables users to create and maintain their digital unique global identities on a decentralized platform and easily show proofs of their identities when needed. Additionally, the users benefit from a faster identity verification process and avoidance of oversharing of data. The report mainly presents the system design of the aforementioned identity management platform. Limitations arising due to privacy concerns and unaffordability of resources have also been discussed. Based on these, further research and development will be conducted.

Acknowledgement

We would like to express our sincerest gratitude to our Final Year Project supervisor, Dr. John Yuen, for his invaluable guidance. We would also like to extend our thanks to the Department of Computer Science, The University of Hong Kong, for providing us with the required resources. Finally, we would like to thank our CAES9542 Techincal English for Computer Science instructor, Mr. Julian Chase, for supporting us through the documentation process of the project.

Contents

List of Figures

Abbreviations

API Application Program Interface

CA Certificate Authority

CLI Command Line Interface

dApp Decentralized Application

DPKI Decentralized Public Key Infrastructure

ION Identity Overlay Network

PKI Public Key Infrastructure

POI Proof of Identity

SSL Secure Sockets Layer

Introduction

The presence of centralized databases has resulted in an increase in the number of data breaches seen in recent times. The hackers are posed with a single point of entry making it easier for them to gain access to users' sensitive data stored on these data-stores. Hence, a decentralized database, where no data is stored at a single location would be a potential solution for this issue.

1.1 Background

Online businesses and web services in the present day are heavily reliant on big corporations such as Google and Facebook for authenticating their users. These businesses often have limited funds and resources which leads them to employ the services of centralized database providers to store their users' sensitive personal identity information. These businesses often use cookies i.e. hidden files on the users' machines which track user behaviour [1]. This data, though trivial and seemingly insignificant is often sold to third-party applications who leverage the data for promoting the sale of their products through targeted marketing strategies. On the contrary, the more relevant personal data collected by these businesses is stored in centralized databases. This sensitive data, though encrypted is vastly susceptible to breach by hackers. The present year itself, has seen numerous data breaches such as the hacking of Citrix, WhatsApp, Instagram, etc. as billions of consumer profiles [2] were vulnerable to identity theft.

The idea of decentralized identity was born as a possible solution that could address the issues posed by the current mechanisms of identity storage and collections. This has seen

vast development as Microsoft [3], and other start-ups like uPort [4] are involved in its development in various sectors like insurance, real estate, financial services, etc. [5]. The decentralized system has been empowered by blockchain. Its features like data immutability and consensus algorithms [6] for decision making, further help the system to be more robust and secure. Furthermore, it allows users to have sole control over their own digital identity, as they are emboldened to share their digital identity for verification purposes.

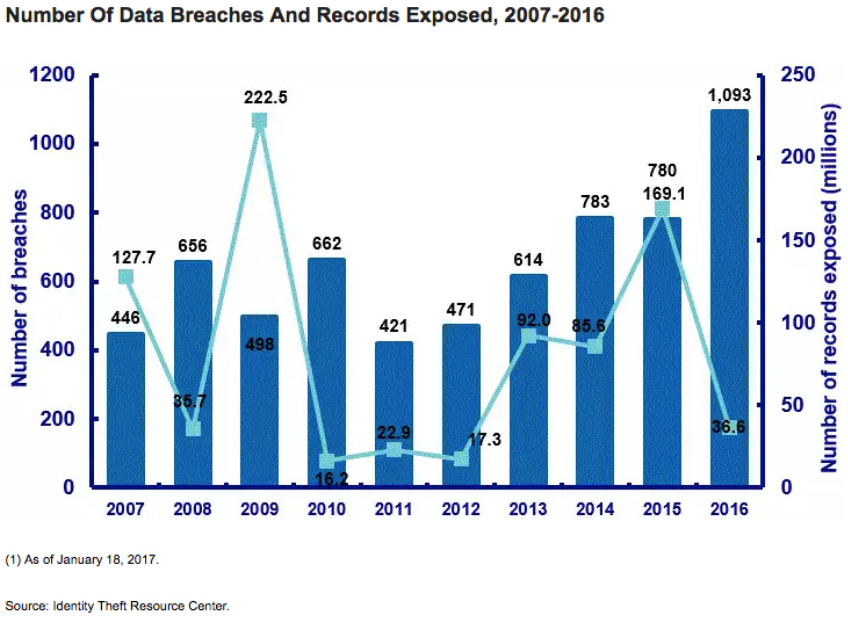

Figure 1.1: Constant Rise in data breaches in recent years

1.2 Objectives

The motivation behind the project is to provide a solution which helps improve the existing standards for identity management and validation. This solution will leverage blockchain to decentralize identity management for people. The premise behind this choice is the empowerment of individuals to have sole control over their digital identities. The platform being created is versatile in the sense that it could be applied in various domains. However, in due course, a relevant application area would be identified such that a comprehensive solution can be implemented which will cater to the more esoteric needs of this domain.

The eventual objective, after the platform has been designed, is the implementation of a platform which allows users to validate their identities. The users would essentially be able to control the data they are sharing through a series of decentralized identifiers which correspond to relevant subsets of the users' identities. This would help prevent oversharing of personal information which is possible in certain situations [7]. The platform would further be extended along the tangent of the SaaS (Software as a Service) model, wherein the platform will be licensed for development based on a subscription model. With this model external organizations could leverage the platform and alter it to be specific to their needs.

1.3 Scope

The project is divided into three parts – (i) a blockchain network to implement the aforementioned features, (ii) construction of decentralized platform to securely store and maintain identity information and (iii) hybrid web application to cater to the individuals' needs for identity verification. The project aims to deliver a well-rounded solution from the view-point of the end-users of this platform. The platform would be able to integrate into other platforms as a feature of the SaaS model.

1.4 Outline of Report

The report briefly outlines the background and motivation behind the project, while offering further insights into the methodologies catered not only for technical but also for non-technical audience. The report will then discuss the current status of the work achieved and scope for future work. Finally, the justifications and limitations of the approaches are included in the methodology section as well. The report concludes by stating the future research directions.

Literature Review

Before diving deep into the project, a thorough research of the fundamental concepts of the technologies involved in the project was carried out. A deep study to understand the nuances of blockchain technology was conducted to determine how it could be leveraged to realize the project's use case. Different types of existing blockchain infrastructures, like Ethereum and Bitcoin, were studied to later aid the decision of choosing a suitable blockchain model for the project. Thereafter, research was conducted on DIDs to develop an intuitive understanding of what digital identities entail and how they could work in the project's context. Moreover, literature review and market analysis was conducted to obtain insights into existing successful projects' way of working.

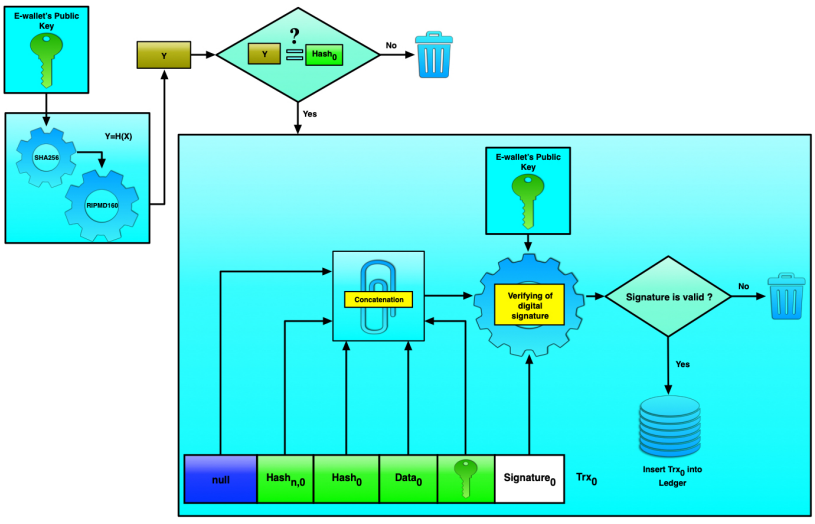

Figure 2.1: Workflow for verifying the Public Key Certificate [8]

Encry's Decentralized Public Key Infrastructure (DPKI) was identified as a good starting point for reviewing the current development and initiatives undertaken in the market to solve

problems similar to our use case [8]. In this growing age of internet security, companies rely on Secure Sockets Layer (SSL) certificates signed by the Certificate Authority (CA) to ensure that the web services being employed by them adhere to the set standards for safeguarding their customers' internet connectivity and sensitive personal data [9]. The presence of multiple signing authorities has forced users' to implicitly trust these CA's who issue digital certificates to maintain the company's Public Key Infrastructure (PKI). A potential solution to this problem is decentralizing the role of Certificate Authorities using the consensus algorithms in the blockchain to assign SSL certificates to companies. The companies simply key in their relevant informant such as proof of business, company domain, etc. for the CA to verify. Upon successful verification, the CA would use its own decentralized identifier to sign off on the certificate [10]. This verification process can further be seen in 2.1. This allows for transparency as any user of the network is simply able to check the votes and certificates assigned by the Certificate Authority.

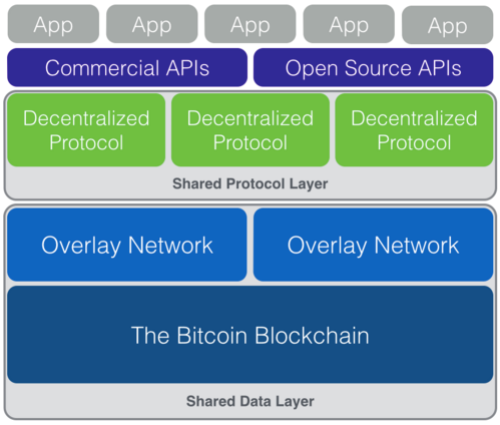

Figure 2.2: The Bitcoin blockchain application stack [11]

It was identified that Microsoft's use case overlapped significantly with this project's [3]. They are basing their solution on the Bitcoin blockchain. Overview of a general Bitcoin blockchain application stack can be seen in Figure 2.2. Microsoft is working on building their own Identity Overlay Network (ION) on top of the Bitcoin blockchain to complete a Shared Data Layer suitable for a DID platform. The ION is able to alleviate the need for

building a new blockchain, even though the primary function of the Bitcoin blockchain is to serve as a ledger for cryptocurrency transactions. It does so by stitching a set of decentralized identifiers into a transaction and then storing the transaction on the blockchain. Since the original blockchain still prevails, there are no changes required to be made to the decentralized protocols of the Shared Protocol Layer. Finally, depending upon how Microsoft wants to market the platform, they will build Application Program Interface (API) for different purposes.

Similarly, uPort is using the Ethereum blockchain to implement the use case. It utilizes the associated state database of the Ethereum blockchain to store the users' data as transactions. While a lot of Microsoft's and uPort's qualitative work is open source, the implementation is proprietary code.

Methodology

Further elaboration on the design choices, algorithms and technologies used for the construction of the blockchain, decentralized platform and identity verification application have been presented below.

3.1 Finalizing Application Domain

The identification of an application area would allow the platform to be extended to solve the solution for a particular domain. A few application areas like a decentralized certification authority for SSL certificates, general identity management platform, decentralized ledger to hasten the bank's KYC were considered to be the potential application area of the platform. Finally, after further research it was identified that the general identity management platform would be the ideal application domain as it can be utilized in nearly every situation. Thus, a realistic sub-domain needs to be identified to make the construction of the platform feasible. Therefore, a simple use case of *proving legal age of drinking at a bar* has been identified as the ideal application to drive the project.

The purpose of this identity management platform is to allow users to create, store, and maintain their own general global identities. The users would be able to validate their identities by registering their personal information on the platform and generating their unique decentralized identifiers. Once the validation has taken place, the users would be able to verify their identities to other parties using their identifiers. For instance, when a person goes to the embassy to obtain a visa, they are required to present numerous documents to verify their identity. This document-intensive procedure can be simplified by the identity

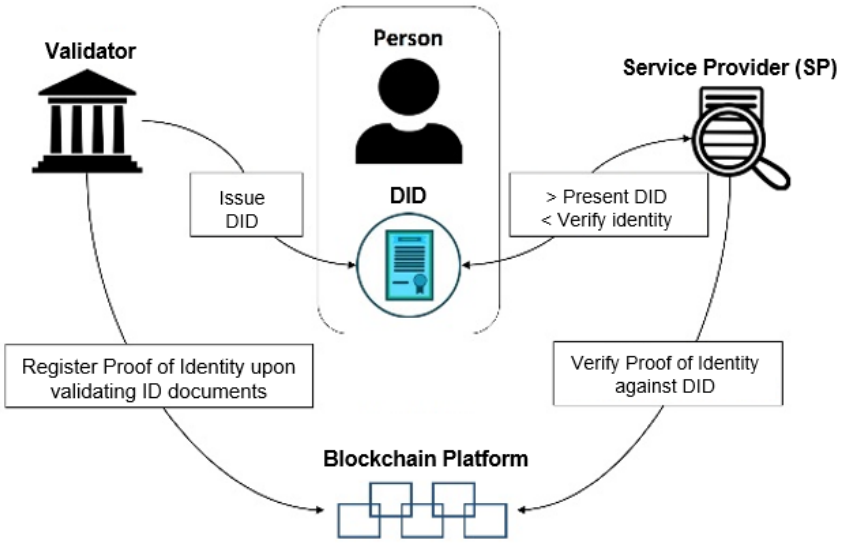

Figure 3.1: A complete use case of the general identity management platform

management platform, requiring the visa official to simply verify the identity of the person stored on the platform against the identifier provided by them. This workflow can be seen in **Figure 3.1**, where the role of service provider is played by the visa official. Furthermore, users could associate different sets of personal data with different unique decentralized identifiers belonging to them to prevent oversharing. For example, a person can link up only their passport information with a unique identifier and provide the flight booking vendor with that particular identifier to verify their identity at the time of flight ticket booking, keeping all other personal information concealed.

3.2 Finalizing the definition of Identity

The identity of an individual involves several details like full name, home address, email address, contact information, government identification records, etc. together which help to uniquely identify the person [12]. Although this generic data may be similar across different domains, the finer details are domain specific. According to Data Protection Principles, the data in question falls under the category of "personal data" since it belongs to a living indi-

vidual. Thus, it is imperative that only relevant data is collected and that it is stored securely [13]. This phase of the project is the most challenging one. The project aims to address this issue of data privacy by ensuring that only the users' proof of identity is maintained on the blockchain. The sensitive personal data documents themselves are encrypted and stored outside the blockchain on the users' personal devices, such that they have sole control over their data [3]. The rationale behind this decision is the need for systems to move away from centralized databases and maintain multiple data stores, so that there exists no one point of data-access. The users' personal devices serve as these data stores in the platform, as they securely house their data.

The process of issuing proofs of identity against personal documents after verification would be carried out by the document issuing authorities. This process would initially be carried out by physical human checking by the registrar. In the future, advanced techniques like Natural Language Processing and Artificial Intelligence can be used to automate this verification process.

3.3 Building the Blockchain

Blockchain technology lies at the base of this project. One of the main design choices was the selection of the Ethereum blockchain over Bitcoin, Hyperledger or a custom-built blockchain. The state database and smart contracts in the Ethereum blockchain would allow the Proof of Identity (POI) to be directly maintained on the blockchain network. In comparison, the Bitcoin and custom-built blockchain do not have these features. Though the Hyperledger network has similar capabilities, it is more suited for private networks. Hence, an Ethereum blockchain would serve the ideal purpose of storing users' POI.

An Ethereum Blockchain consists of four trie's namely the transaction, state, receipts and storage trie (Refer to Figure 4.2). The transaction trie as the name suggests stores the transactions of the blockchain. Similarly, the storage trie is where the data from the smart

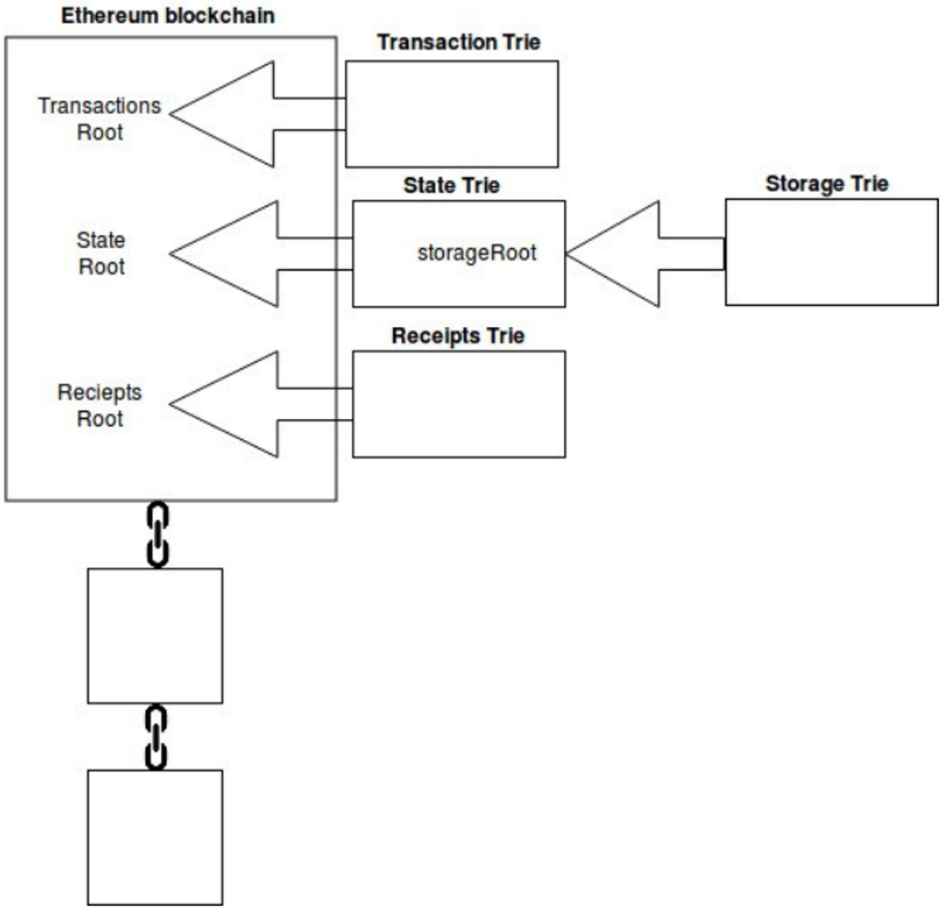

Figure 3.2: System Design of an Ethereum Blockchain

contracts live. They are associated with every account on the blockchain network. The data in the trie is encrypted using the account's public key to maintain security of the data. The combination of all storage trie's is the state trie. It 's ephemeral nature makes it suitable to store data such as the wallet balance, as they increase/ decrease with every transaction. The roots in the Ethereum blockchain are a hash of their respective trie's which means that if any data in the trie's is tampered the hash changes. This implies that the data in a blockchain is tamper-proof, and that data can only be added through consensus.

A challenge will be to allow amendments to be made to the identity data on the blockchain, since blockchain data is theoretically immutable. This is where the smart contracts on the

Ethereum blockchain network are beneficial. Smart contracts are computer programs within the blockchain that automatically enforce the obligations defined in the contract when its criteria has been satisfied. For instance, the POI for an individual to prove whether he is of legal drinking age, was registered two weeks before he turns legal. In this case, a smart contract could be defined that would modify this information on the POI. As mentioned earlier, this data is stored in the state trie of the blockchain which is ephemeral in nature. However, this capability is not necessary for our use case and could be implemented if the project is taken ahead.

3.4 Identity Verification Application

A hybrid (platform independent) web application will be created to serve as a means for users to interact with the DID platform. Users will be able to securely store their personal identification documents on their devices and be able to register themselves on the platform via the proof of their identities on the blockchain. A functionality to create more than one unique global identifiers will be provided to the users so that they can link up the different sets of details with different identifiers. The application will entail both sides of the verification process - the identity prover side and the identity verifier side. The complete DID system (the DID blockchain platform and the end-user application) aims to enable holistic use cases like those discussed in Section **3.1** to be successfully realized.

The challenge to achieve platform independence i.e. the ability to access the application through various devices (laptop, tablet, smartphone, etc.) but at the same time, also leverage a decentralized database i.e. not using a single database for multiple devices, will be solved by building an Ethereum Decentralized Application (dApp) using React-Native. React-Native is a cross-platform application development framework (for developing applications for both Android and iOS Operating Systems). The framework has modules, which allow for the easy integration with *Ganache*, a development environment which allows the blockchain network to be simulated in a localhost environment [14] (deploy software on local machine without

integrating with the cloud infrastructure for ease of testing and development). **Figure 4.2** shows the simulation of the Ethereum blockchain using the Ganache Command Line Interface (CLI).

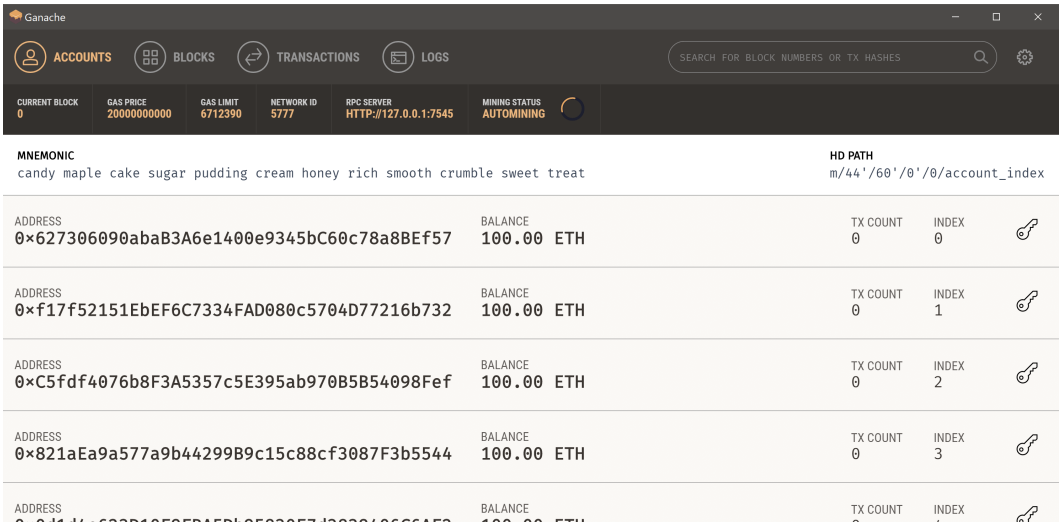

Figure 3.3: Simulation of Ganache CLI
[14]

Taking the use case of an individual proving he is of legal drinking age at a bar to explain the workflow of the application. The steps are as follows:

1. The user first creates an account on the dApp.
2. Registers his bio-metrics on the dApp.
3. Subsequently the user takes his documents to the issuing authority where the registrar confirms the validity of the document and issues the POI.
4. A transaction is registered against the user's account for a small transaction fee. This allows a wallet to be attached to the account.
5. The registrar signs the POI and puts it onto the blockchain.
6. The user can use the DID registered against this POI to validate his identity.

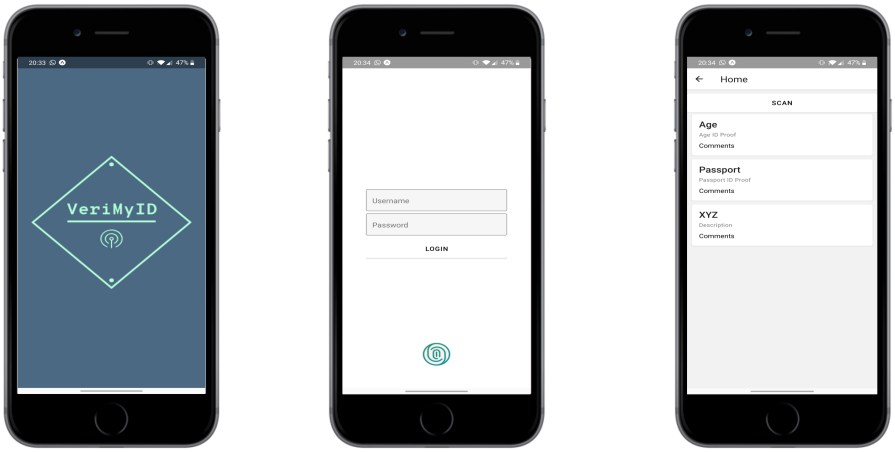

Figure 3.4: Identity Management Application - Prover

7. The verifier scans the QR code, which contains the hash of the user's public key, registrar's public key, DID and the time of creation.

8. The verifier can then access the data stored in the POI to verify the user is of legal drinking age.

Figure 3.5: Identity Management Application - Verifier

Project Plan

The sections below discuss the current progress and the work to be done in the next iteration.

4.1 Identification of Application Area

A decentralized identity platform is multifaceted, as it may be modified to be of pertinence to multiple domains. The identification of a domain is necessary to address the more esoteric needs of the platform. Although various application domains were discussed in section **3.1**, it was finally decided that a general identity management platform would be the best fit for our use case.

4.2 Literature Review

Over the past few months, considerable progress has been made as extensive literature review has been undertaken to analyze existing work being done in this domain. Three companies namely Encry's DPKI, Microsoft and uPort have been identified to have implementation similar to our use case. Although, Microsoft and uPort are hands-on with the development of a decentralized identity platform, their actual implementation is proprietary.

4.3 System Design

The literature review conducted was helpful in finalizing the system design for the platform. As seen in Figure 4.1 the system design of the platform consists of 3 parts:-

1. Ethereum Blockchain Network

2. Smart Contracts which communicate with the blockchain to store users' POI on the state database.

3. An Ethereum dApp built using React-Native to assist the prover to verify his identity and the verifier to verify the corresponding POI.

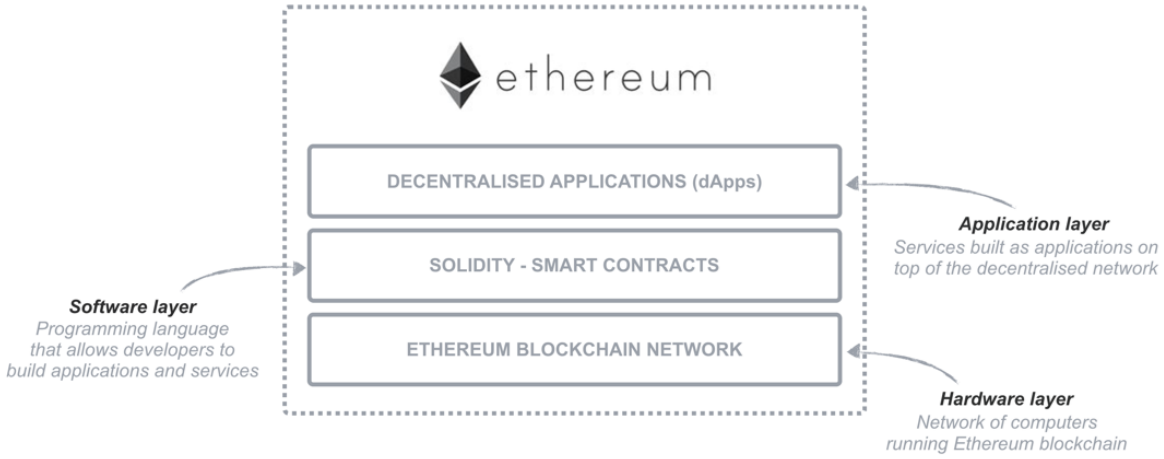

Figure 4.1: System Design of our use case

4.4 Implementation of the Idea

A cross-platform Ethereum dApp for the purpose of identity verification as mentioned in **section 3.4** has been initiated. The data displayed on the dApp is just a placeholder and would later be associated with the user's account on the state database of the blockchain.

Furthermore, the implementation of blockchain networks on Ethereum has been initiated. Figure 4.2 shows the blocks on the Ethereum blockchain being synced before further development may be conducted.

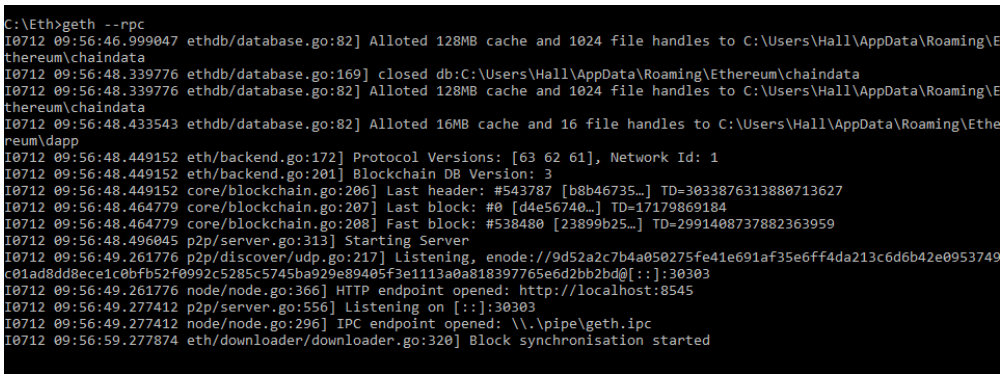

Figure 4.2: Implementation of Ethereum Network - Sync transactions

Finally, a website has been created for the project so that viewers can have an idea of what the project entails, current progress and development, as well as the timeline for project progress. The link for the website is at: **https://i.cs.hku.hk/fyp/2019/fyp19055/**

4.5 Scope of Future Work

The implementation of the project is as per the deadlines defined in the *Product Plan* on the website. In the final iteration of the project, the actual implementation details as defined in **section 4.3** would be developed. The first step in the process would be the identification of how the data would actually be stored in the Ethereum blockchain using the smart-contract objects developed using Solidity.

Once the blockchain network has been finalized, the focus would shift onto the development of API which could communicate the data between the blockchain and the dApp. Additionally, research would be conducted to finalize how the issuing authorities would register the POI with the user's account once the documents have been successfully verified.

Conclusion

The report presents an efficient and secure digital identity validation solution using blockchain technology. The solution has two components - the DID platform on the blockchain and the end-user verification application for both prover and verifier. Users benefit from this decentralized solution by exercising absolute control over their digital identities and eliminating their dependencies on centralized authorities. The verification process is made much simpler and efficient. Oversharing of data is also avoided.

It has been found that a lot of work is being done in this area involving many different application domains. Microsoft and uPort, in particular, are implementing a solution that is very similar to this project's vision. The main application idea identified for the project is that of a general identity management platform, but with a more specific use case of 'proving legal age of drinking at a bar'. The system architecture has been planned in a manner that allows users' personal data documents to be securely stored only on their personal devices while using the blockchain decentralized platform to store their proofs of identity. These proofs of identity will suffice for a prover to prove their identity to a verifier. Instead of a custom-built blockchain infrastructure that was being considered previosuly, the decentralized identity platform will implement an Ethereum blockchain. The verification application is a cross-platform hybrid web application developed using React-Native. All the aforementioned recommendations will be assessed further and amended, if required, as the project progresses.

Limitations and challenges, as mentioned in each of the preceding sections and subsections, will be addressed through appropriate research and development.

References

[1] L. Forensics. (2019). Cookies – what are they and what are the rules for business websites using them? [Online]. Available: https://www.leadforensics.com/what-are-cookies-and-can-businesses-use-them/. [Accessed: October 20, 2019].

[2] S. Turner. (2019). Recent data breach roundup: May 2019, [Online]. Available: https://www.identityforce.com/blog/data-breach/recent-data-breaches-may-2019. [Accessed: October 20, 2019].

[3] Microsoft. (2018). Decentralized identity, [Online]. Available: http://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE2DjfY. [Accessed: October 20, 2019].

[4] Kames. (2019). The basics of decentralized identity, [Online]. Available: https://medium.com/uport/the-basics-of-decentralized-identity-d1ff01f15df1. [Accessed: October 20, 2019].

[5] B. Technologies. (2019). Blockchain applications in identity, [Online]. Available: https://www.blockchaintechnologies.com/applications/identity/. [Accessed: October 20, 2019].

[6] H. Anwar. (2019). 6 key blockchain features you need to know about! [Online]. Available: https://101blockchains.com/introduction-to-blockchain-features/. [Accessed: October 20, 2019].

[7] C. Grundy. (2019). Why decentralized identifiers are changing the future of the internet, identity and finance, [Online]. Available: http://selfkey.org/decentralized-identifiers-article/. [Accessed: October 20, 2019].

[8] R. Nekrasov. (2017). Dpki: Addressing the disadvantages of centralized pki by means of blockchain, [Online]. Available: https://habr.com/en/company/encry/blog/461731/. [Accessed: February 15, 2020].

[9] GlobalSign. (2017). What is ssl? [Online]. Available: https://www.globalsign.com/en/ssl-information-center/what-is-ssl/. [Accessed: October 20, 2019].

[10] L. Hentschker. (2018). Decert: A decentralized certificate authority, [Online]. Available: https://www.boazbarak.org/cs127/Projects/decert-FINAL.pdf. [Accessed: October 20, 2019].

[11] in. (2018). Blockchain app stack, [Online]. Available: http://blog.snap.hr/18/01/2018/how-to-get-a-job-in-crypto/blockchain-app-stack/. [Accessed: December 5, 2019].

[12] K. L. Michael Sweeney. (2018). What is pii, non-pii, and personal data? [Online]. Available: http://piwik.pro/blog/what-is-pii-personal-data/. [Accessed: October 20, 2019].

[13] H. T. L. Joshua Cole. (2019). Hong kong: Data protection 2019, [Online]. Available: http://iclg.com/practice-areas/data-protection-laws-and-regulations/hong-kong. [Accessed: October 20, 2019].

[14] Hackernoon. (2018). Bringing the blockchain to react native, [Online]. Available: https://hackernoon.com/bringing-the-blockchain-to-react-native-98b76e15d44d. [Accessed: December 5, 2019].