香 港 大 學

THE UNIVERSITY OF HONG KONG

Decentralized Identity Platform by Blockchain

Project Plan

Ankit Tibrewal, Shivansh Mittal

Department of Computer Science

Supervisor: Dr. John Yuen

September 29, 2019

Contents

Background

At present, most online businesses and web services depend upon big corporations to carry out user authentication and verification processes. This dependency is established as a consequence of smaller businesses lacking the appropriate resources required to safely store the users personal information and private credentials [1]. Therefore, the task of securely storing all their customers sensitive data is outsourced to the huge centralized databases maintained by companies like Google, Microsoft, Facebook, etc. As a result, the users have to put the security of their personal data at stake to be able to participate with the web services offered by these small businesses.

On one hand, the data collected from users could constitute of seemingly trivial and inconsequential data such as their speed of scrolling on a webpage or their browser window size [2]. Big corporations often trade this data [3] with other parties looking to leverage this data to improve their operations and sales, for example, through ad-targeting [4]. On the other hand, a lot of the personal data collected could be significant data that is crucial to a users identity, for instance, passport number or HKID number. Since all this information is stored in a centralized database, although encrypted and safeguarded, hackers are required to get past only one point of failure to cause a data breach. Few of the biggest data breaches of recent times such as hacking of Yahoo!, Facebook, and Marriott International resulted in millions of digital profiles becoming susceptible to identity theft [5].

The inception of the concept of decentralized identity took place, inspired by the need to address these identity related issues. Alongside large firms like Microsoft, many startups have been working on solutions involving decentralized identities in multivarious industry domains. Some applications include digitalization of passports for tenants and landlords in house rental markets, mitigation of fraud in medical prescriptions and insurance claims, building trustworthy data sharing platforms, private credential storing tools, etc. [6]. Blockchain serves as the foundation layer for the decentralized system which has the decentralized identities rooted in it. This empowers the users to enjoy self-creation, self-ownership, and absolute control over their global unique digital identities. Moreover, owing to blockchains most crucial characteristic of immutability, high level of security and robustness of the system is ensured [7].

Objectives

This project aims to deliver an improved alternative to existing digital identity validation mechanisms. This alternate solution will leverage blockchain to create decentralized (not relying on centralized databases for storage of data [8]) identities for people, thereby handing them absolute control over their digital identities and eliminating dependencies on centralized authorities. Consequently, it can be hoped that misuse of peoples personal data by centralized parties will be abated and cases of identity theft caused by data breaches can be mitigated.

The decentralized identity solution can be applied to many different industrial domains, as discussed in the background. Initially, a particular application area will be identified and an appropriate digital identity will be defined in that context. The decentralized identity platform will aim to cater to different use cases in that area. The identity will not be limited to fundamental parameters such as name, age, date of birth, etc., but will also encompass more esoteric application specific details. Moreover, the platform will include the functionality allowing users to not only maintain their identities, but append information to it.

The ultimate goal, after the blockchain platform is set up, will be to deliver an application to make the identity verification process quicker and more efficient. Users will be able to exercise control over the extent to which they wish to share their personal details and thus, oversharing of data will be avoided [1].

While the scope of the initial project plan covers the aforementioned, the project envisions to provide a general domain-independent SaaS (Software as a Service) implementation of the decentralized identity platform, allowing other organizations to mould the solution to address their particular needs.

Methodology

The project involves creation of a blockchain and the decentralized identity platform, followed by development of an application to address users needs for identity verification in a particular domain. The different tasks have been discussed below.

3.1 Identifying Application Area

A decentralized identity platform is multifaceted as it may be modified to be of pertinence to multiple domains. The major application domains that have been identified so far are:

- **An efficient solution to the bank's Know Your Customer policy-** Banks, on a daily basis, ask their customers to provide them with basic personal information such as their name, date of birth, tax returns, etc. so that the banks can validate their track record and offer them services such as opening up a bank account, loan applications, remittance of money to international banks, etc. This procedure comes under the Know Your Customer (KYC) policy, allowing the banks to assess customer credibility, thereby helping them to prevent cases of defaults and money-laundering [9]. This is often a cumbersome process for both bank employees and customers, involving a lot of paperwork. The whole process can be made much more efficient by maintaining digital identities of customers, which consists of their basic information, bank transactions, etc., on the blockchain. For future reference, banks can simply verify the digital identity of a customer and provide relevant services to them thereafter.
- **A General Identity Management Platform-** This platform would 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 management platform, requiring the visa official to simply verify the identity of the person stored on the platform against the identifier provided by them. 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.

As the project moves forward, further research will be conducted to identify a suitable application area. Final selection would be made either from one of the aforementioned ideas,

or from a different new domain.

3.2 Defining 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 [10]. Although this generic data may be similar across different domains, the finer details are domain specific. Once the application domain has been finalized, the constituents of identity in that domain would be identified and thereafter, a firm definition of identify would be established for the project.

According to Data Protection Principles, the data in question falls under the category of personal data since it belongs to a living individual. Thus, it is imperative that only relevant data is collected and that it is stored securely [11]. 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 [7]. 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.

3.3 Setting up the Blockchain

Blockchain technology lies at the base of this project. One of the main design choices considered is that of a custom-built blockchain. This would allow much more freedom in implementing the solution and the platform will not be required to adapt to the limitations of other popular choices of blockchain infrastructure such as Ethereum and Hyperledger, which are quite advanced and designed for intensive data-centric applications. They consist of features which are superfluous to the requirements of this project. Hence, a simple tailor-made blockchain would serve the purpose of storing the users proof of identity.

3.4 Application for Identity Verification

A hybrid web application will be created to serve as a means for users to interact with the decentralized identity 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 processthe identity prover side and the identity verifier side.

Schedule And Milestones

4.1 September 2019

i. Research on technology behind blockchain and decentralized identity platform

ii. Literature review and market analysis of existing work on decentralized identity

iii. Creation of detailed project plan to outline project background, objectives, methodology, schedule and milestones

iv. Website development to display key information about the project

v. **Phase 1 Deliverables due 29th September, 2019**

4.2 October 2019 to December 2019

i. Identification of application area for the decentralized identity platform

ii. Literature review on existing work, competitor analysis, and research on existing successful solutions in the identified application area

iii. Testing of different types of blockchain infrastructures (Eg. Ethereum, Hyperledger) to identify suitable system architecture

iv. Finalization of definition of identity through identification of constituent personal user data fields

v. Analysis of standards for data collection, privacy, and storage security

4.3 January 2020 to February 2020

i. Preliminary implementation of blockchain and decentralized identity platform

ii. Initiation of construction of the application for identity verification

iii. Creation of a detailed Interim Report, explaining current progress, implementation, and scope of development

iv. Initial Presentation Deck, for presentation of current progress and other analysis

v. **First Presentation between 13th to 17th January 2020**

vi. **Phase 2 Deliverables due 2nd February 2020**

4.4 March 2020 to April 2020

i. Testing and finalization of the decentralized identity platform and the identity verification application

ii. Creation of Final Report to comprehensively explain the work done, problems solved, limitations, and future scope

iii. Final Pitch Deck to demonstrate the use of the platform and present any significant findings

iv. **Phase 3 Deliverables due 19th April 2020**

v. **Final Presentation between 20th to 24th April 2020**

4.5 May 2020 to June 2020

i. **Project Competition on 5th May, 2020**

ii. **Project Exhibition for selected projects on 3rd June, 2020**

References

[1] 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: 25.09.2019).

[2] M. Hanif. (2018). What data is collected about you online and how to stop it, [Online]. Available: http://www.globalsign.com/en/blog/what-data-is-collected-about-you-online/. (Accessed: 25.09.2019).

[3] L. Matsakis. (2019). The wired guide to your personal data (and who is using it), [Online]. Available: http://www.wired.com/story/wired-guide-personal-data-collection/. (Accessed: 25.09.2019).

[4] C. Hoffman. (2016). The many ways websites track you online, [Online]. Available: http://www.howtogeek.com/115483/htg-explains-learn-how-websites-are-tracking-you-online/. (Accessed: 25.09.2019).

[5] K. Kiesnoski. (2019). 5 of the biggest data breaches ever, [Online]. Available: http://www.cnbc.com/2019/07/30/five-of-the-biggest-data-breaches-ever.html. (Accessed: 25.09.2019).

[6] H. Garneau. (2019). Four decentralized identity startups selected from around the world for intensive 12-week incubator, [Online]. Available: http://www.globenewswire.com/news-release/2019/08/28/1907809/0/en/Four-decentralized-identity-startups-selected-from-around-the-world-for-intensive-12-week-Incubator.html. (Accessed: 25.09.2019).

[7] Microsoft. (2018). Decentralized identity, [Online]. Available: http://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE2DjfY. (Accessed: 25.09.2019).

[8] A. Tar. (2017). Decentralized and distributed databases, explained, [Online]. Available: http://cointelegraph.com/explained/decentralized-and-distributed-databases-explained. (Accessed: 25.09.2019).

[9] vikaspedia. (2016). Know your customer guidelines, [Online]. Available: http://vikaspedia.in/social-welfare/financial-inclusion/know-your-customer-guidelines. (Accessed: 25.09.2019).

[10] 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: 25.09.2019).

[11] 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: 25.09.2019).