Project overview and plan - A Blockchain-Based Credential Store

Supervisor: Richard Banach
Huanjie Guo
University of Manchester
huanjie.guo@postgrad.manchester.ac.uk

I.INTRODUCTION

Since the network was invented, people store their information in one database, making the system vulnerable to disasters. It is often reported that some staff wrongly operate the system, and then all data was deleted, or someone secretly changes the data in the database. Nowadays, students need to request their degree when they leave their schools and go to secure a job in society, and a degree certificate can be one of the most important personal information in our life. Therefore, data like a degree certificate should have a reliable mechanism to keep it safe. This mechanism should record every modification on the data, which is safer because we can track the process in the future. To store our data in a single database is not a good idea since data may lose or be tampered with by offenders.

With the rapid development of blockchain, a growing number of applications are designed to run on the blockchain. Companies choose to put data on the blockchain because data could have a backup in every node, and people can track every transaction and operation in the real world. As more and more ideas based on blockchain break out after its invention, this new technology seems to push the Internet into a new stage and reshape our society gradually.

To solve the problem of the traditional centralised degree management system, this project aims to take advantage of blockchain and smart contract to design a decentralised degree management system base on Ethereum. In this system, digital degree certificates are encrypted and stored in every computer that serves as a node that runs Ethereum virtual machine worldwide. The system includes degree issue, degree search, degree revoking and degree verification.

II.BACKGROUND

Lamport, Shostak and Pease (1982) published The Byzantine Generals Problem. They discussed a problem when some traitors in the troops would provide fault information and how could the generals make the decision. For a distributed system, it is essential to solving conflicting information in different parts of the system. When data is stored in different computers and one of them changes data on it, should other computers change the data correspondingly, or they should stay the same and force the changed one to roll back. During these several decades, many consensus algorithms have been issued to solve the Byzantine generals problem.

Nakamoto (2018) designed a peer-to-peer electronic cash system. In this system, the terminology of "blockchain" was invented, which means that transaction information is added into the block, and the next block will contain the hashcode of the last block. The private key is everything in the bitcoin system since it can calculate the public key and then use a one-way hash function to get the address. Users can transfer token to each other with their private key. Owning the private key can prove that this person is the owner of the wallet. What is more, a private key has 256 bits, which means that it is computationally impossible to guess someone's private key. The most innovative thing is that every personal computer in the world can use the computing power of the CPU to compete for the right of keeping accounts. Data of transaction are stored in every node that runs the Bitcoin program, making data more transparent and undeniable. Besides, Satoshi mentioned Proof-of-Work in his white paper, which solves the problem of majority decision making. If the system runs based on one-IP-one-vote, it will lead to a problem that someone can use many IPs and always get more chance to make the decision. PoW is like one-CPU-one-vote, and only when people use much electric power will they have a higher possibility of getting the right to write the data into the blockchain. Bitcoin is creative and innovative, and it is a successful combination of cryptography and finance in our modern society.

Although proof of work can make the network more secure and reliable, it needs to consume lots of energy, and people want to find a more efficient consensus algorithm that can save computing power. Kiayias et al. (2017) modified the consensus protocol of Bitcoin, replace Proof-of-Work with Proof-of-Stake. PoS means a form of ownership of the token, calculating the weight of a node based on how

many currencies it holds and not computing power. However, the security of PoS still needs to be tested in the future. It is difficult for designers to find a more suitable and scientific algorithm to generate and allocate the token when initialising the blockchain.

Vitalik Buterin (2014) published a white paper of Ethereum. Unlike Bitcoin, Ethereum is designed to serve as a blockchain with a built-in fully fledged Turing-complete programming language. A programmer can write "contracts" that can encode arbitrary state transition functions, giving a tool for people to design the program they want. Smart contract in Ethereum acts as a box and only can be opened when the criteria are met. When the code in smart contracts is translated into a binary executable file, it was then deployed on the blockchain. After that, when users call smart contracts by interfaces, Ethereum virtual machine will execute the code of smart contracts and then modify the state of data on the chain. Once data was changed, it will notify every node to change and ensure the running nodes can keep consistent. The invention of smart contract facilitates the development of the blockchain and makes the system more secure and reliable.

Gresch *et al.* (2019) designed a reliable diplomas verification system based on blockchain and smart contracts. They pointed out that there are two ways to implement a diploma system based on blockchain. The first one is to store the data directly into the blockchain, and everyone can validate it by data on the chain. Besides, people can track the operation that happened on the blockchain, which is more transparent. However, it has its weakness that personal information can be leaked and be used for some illegal purposes. The second way is to store students' diploma into the database of universities and record the hash value of their certificates into the blockchain. When a company receives diplomas, it can calculate its hash value and verify it by the blockchain. Recording hash values into the blockchain solves the problem of privacy, but if the diploma information in universities' database is lost, the hash on the blockchain becomes useless. Using which mechanism to store data on the blockchain is totally depends on which situation it is.

III.RESEARCH METHODLOGY

Figure 1 illustrates the academic degree certificate flow based on Ethereum. It shows the

process of degree search, degree verification, degree issue and degree revoke.

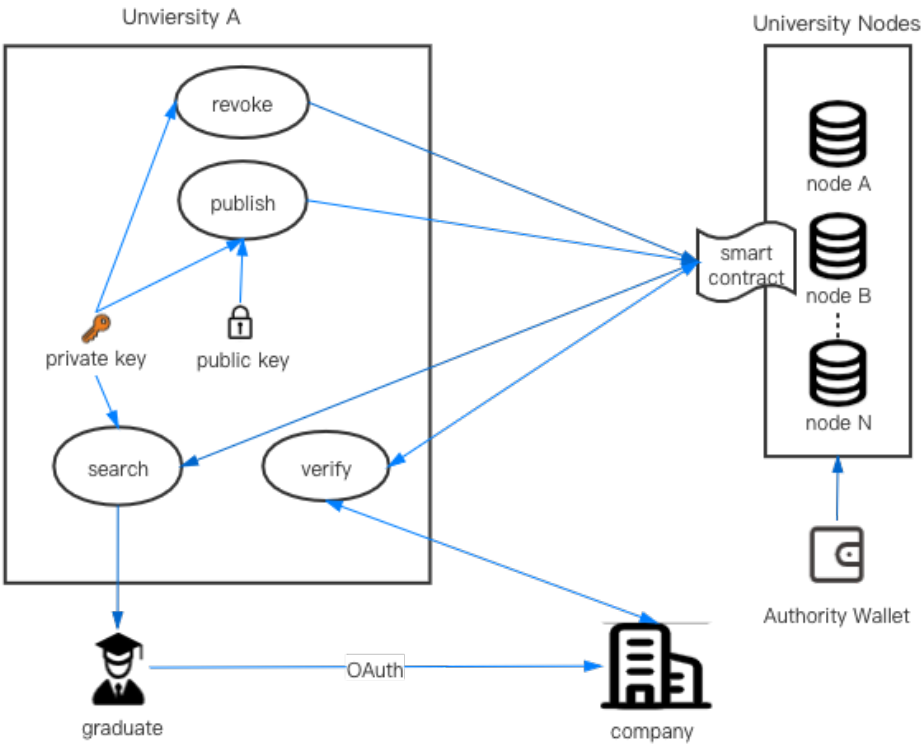

Figure 1 Degree certificate flow based on Ethereum

In this system, our blockchain will initiate 1000 tokens into the authority wallet, which the government controls, and the amount will not increase anymore in the future. Let us call this token "University Identification Coin" - UIC. When a university wants to write their graduate's information into Ethereum, this university needs to request the authority. The authority will send 1 UIC to the university after verifying its identity. After that, this university can run as a node, and it gets the right to append data on the blockchain. In this blockchain, only those universities that have more than 1 UIC can call the smart contract. If one university tries to make a mess in this blockchain, there will be a smart contract to cancel the right of universities. This cancellation mechanism can run on one-UIC-one-vote. If there are more than 666 UICs voted for cancelling the node of the evil university, 1 UIC in the address of this university will automatically transfer to the

authority address. After that, this node cannot keep accounts anymore.

For every university, they should run a node and make the chain more reliable and resilient. Each university has its own private key, and they can calculate their public keys and addresses by private keys. Universities need to offer their address to the authority, and then people can search online and find that who is the owner of this address. The public key here is to generate the address and be applied to encrypt student's personal information. Also, the private key of a university is essential because it can be used to do the signature then tell everyone that the sender is the holder of this address.

When a student graduates, universities will generate a digital degree certificate that has a specific number. After that, universities use RSA to encrypt the certificate c and get E(c). Then universities use the MD5 algorithm to get the digest of the certificate D(c). Finally, universities put E(c), D(c) and the certificate number into a data structure and upload it to the blockchain. Although some systems use plaintext, uploading the raw data without encryption is not a good choice because everyone can get the data from the blockchain, and personal information will be leaked.

If universities want to revoke degree certificates that they have issued before, what universities should do is search the student's name in its database and get his certificate number, then call the smart contract of Ethereum, which is designed for cancellation with the certificate number. After that, the blockchain will check if this university has issued this certificate before and do the corresponding operation immediately.

Students can request their own degree certificates from the university, and universities will search data from the blockchain. As we know, data is encrypted on the blockchain, and universities need to use the private key to decrypt it then return it to students. It is so convenient for students because although data is encrypted in the blockchain, students can get their certificates directly from their universities. On the other hand, criminals cannot get the original certificates as these data have been encrypted.

Besides, Students can authorise companies to verify their digital degree certificate. Companies will send students' digital degree certificates to their universities. Then universities calculate its MD5, compare it with the one stored on the blockchain and return the

result to the company. If companies do not trust universities, employers can calculate the MD5 of certificates themselves and then check it with the MD5 on the blockchain.

Ⅳ.ETHICS AND PROFESSIONAL CONSIDERATIONS

Data Ethics: Blockchain is an open-source ledger, and everyone on the Internet can get access to data on the blockchain. Therefore, it brings some privacy ethics problems that users should know that their personal information will be written on the blockchain on which everyone can search. Although this system can use cryptographic technologies to encrypt some personal data, some data should not be encrypted because it can be used to search on the blockchain. Besides, although some encryption algorithms like RSA and AES are used widely and regarded as secured enough, with the development of CPU's performance, these cyphertexts may be decrypted by brute force to search the keyspace in the coming future. Therefore, it is necessary to inform graduates how their data will be used on the blockchain and get their permissions in advance.

Smart Contract Ethics: A smart contract is a program, and when the specific condition is met, it will be executed automatically. Smart contract tends to build a new society based on peer-to-peer, rather than peer-to-government. We can build our trust system on smart contracts then Ethereum virtual machine will execute all rules. In this system, we do not need to rely on individuals or the government to supervise since every rule can be written on smart contracts. In a world built by smart contracts, algorithms will finally replace laws. With the rapid development of smart contract, a new social contract will fulfil human potential (Tang *et al.*, 2019). When this period is coming, human will go into a new era in which we live together with computing power.

Decentralization Ethics. Ethereum requires every node to take part in mining, and some of these nodes will do the same calculation, which is a waste of energy. Government, companies, and most of us upload information to cloud computing providers like AWS and Google Cloud, reducing costs. On the other hand, surrendering our data will accumulate the power of providers, and in the future, it could be a massive risk for our human being (Tang *et al.*, 2019). Without a good understanding of both the benefit and drawbacks of blockchain, it

is hard for a human to decide when to choose decentralisation and choose centralisation.

V.RISK CONSIDERATION

Key Storage: Private keys of universities might leak. Once the private key is leaked, the offender who gets the key can have the right to write something on the blockchain. It could be used to forge degree certificates and revoke someone's certificate maliciously. It is one of the most important for universities to keep them private key safely.

51% Attack: There is a risk that if the system runs on Proof of Work. It may suffer from 51% attack. The attackers might control more than 51% hashing power of this network, and then they can modify data on the blockchain. This system needs every university to take part in and contribute their computing power to the network and make it run reliably and stably.

Operation Risk: When people use centralized systems and store some data wrong on the database, they can modify it. However, it is difficult to change the data that have been written, and operations on the blockchain could be irreversible. Before operating smart contracts, users need to know what this smart contract will do actually. Otherwise, it might bring a huge inconvenience to the user. Every node needs to have a mechanism to avoid some wrong operations too. On the other hand, some smart contracts for reversing the operations should be designed to protect users.

VI.PROJECT EVALUATION

Blockchain Testing: This blockchain will be evaluated to see if it is robust enough. We will build up three nodes to run on the same private network and test if the whole decentralised system can work stably. Besides, the test will include calling the smart contracts that written on the blockchain and observing that if the Ethereum virtual machine will execute the code correctly.

API Testing: There is a back-end system in universities used for students and companies to call the interfaces. We would use different parameters to test the system and see if it can return what we want. What is more, the design code of interfaces should be checked since it will link to the database of universities.

Penetration Testing: Private keys that control universities' wallets are stored in the server of universities, so the server must be safe enough to prevent the node from being exploited by hackers. We will use NMap to scan the ports of

the server and see if only necessary ports are open in case that hackers can get the root account through system vulnerabilities and control the sever of universities.

Ⅶ.PLANNING

Figure 2 shows the project planning.

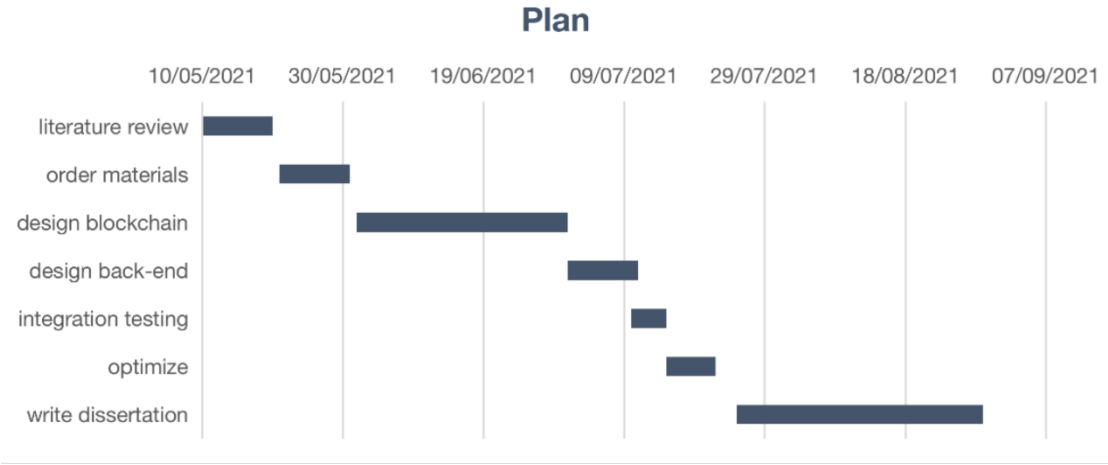

Figure 2 Project planning

REFERENCES

Buterin, V. (2014) "A next-generation smart contract and decentralized application platform," *Etherum*, (January), pp. 1–36. Available at: http://buyxpr.com/build/pdfs/EthereumWhitePaper.pdf.

Gresch, J. *et al.* (2019) "The proposal of a blockchain-based architecture for transparent certificate handling," *Lecture Notes in Business Information Processing*, 339(July), pp. 185–196. doi: 10.1007/978-3-030-04849-5_16.

Kiayias, A. *et al.* (2017) "PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake," *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS'16*, 1919(January), pp. 1–27. Available at: http://peerco.in/assets/paper/peercoin-paper.pdf%0Ahttp://fc17.ifca.ai/preproceedings/paper_73.pdf%0Ahttp://arxiv.org/abs/1606.06530%0Ahttps://papers.ssrn.com/sol3/papers.cfm?abstract_id=2977811%0Ahttp://dl.acm.org/citation.cfm?doid=2976749.2978389%0Ahttp.

Lamport, L., Shostak, R. and Pease, M. (1982) "The Byzantine Generals Problem," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 4(3), pp. 382–401. doi: 10.1145/357172.357176.

Nakamoto, S. (no date) *Bitcoin: A Peer-to-Peer Electronic Cash System*. Available at:

www.bitcoin.org (Accessed: May 5, 2021).

Tang, Y. *et al.* (2019) “Blockchain ethics research: A conceptual model,” *SIGMIS-CPR 2019 - Proceedings of the 2019 Computers and People Research Conference*, pp. 43–49. doi: 10.1145/3322385.3322397.