UNIVERSITY OF HONG KONG

Computing Statistics on Encrypted Data With Blockchain

Project Plan

Researched By

Mengna Su

Supervised By

Dr. S.M. Yiu

Sep 2019

Contents

1 Introduction

Nowadays as we step into the big data era, data collection and usage is of more importance. However, the security and privacy of data are not guaranteed with a centralized server. One may heard of the data leak from Facebook in which many Facebook users' data was collected by an unauthorized third party when users participated in a questionnaire with their Facebook account.

A solution to such problems is to use Blockchain. Blockchain could be regarded as a decentralized database[1], in which data or transactions are stored in different blocks, and updated by appending the updates to a new block. Blockchain secured its data by reaching a distributed consensus. In this way, each user of the Blockchain will get a copy of the data stored. Thus, the system is robust as hacking the system means to cheat at least half of all users using Blockchain, which is hardly possible.

This project aims to provide a solution to collect and analyze data securely through Blockchain. The users will be providing their encrypted personal data to Blockchain, miners on the Blockchain will collect and analyze those data, and the data requesters will finally get the statistical result (e.g. the sum, average of the data) from the Blockchain. In this way, the personal data from data providers could be guaranteed secure and private, as the data requester with decryption key is not able to access those personal information, while the miners are not able to interpret the encrypted data without the decryption key.

2 Background

This section is going to provide the backgrounds for the project, including Blockchain technology, Ethereum platform and encryption mechanism, which are the key ideas of the project.

2.1 Blockchain

Blockchain technology is essentially an open and distributed ledger which could record transactions between different parties efficiently and securely[2].

Thus it's often used in cryptocurrency, and a well-known implementation is Bitcoin.

Besides the normal Blockchain users who can access and update the information on the chain, there is another very important role in the Blockchain called miner. A key idea of Blockchain is to reach an agreement on the updated information among all users of Blockchain, and miners are mainly responsible for this purpose. Miners in the Blockchain will collect the updated data among a certain period of time, verify those data and append them on Blockchain. As many miners will verify the same data during the period, they will compete to broadcast the collected information to all Blockchain users through some protocols, so as to finish the data update process. From this perspective, it's clear that the honesty of the miners are a big issue for the security of Blockchain, which is also a key problem of this project.

2.2 Ethereum Platform

There are many implementations of Blockchain, one very famous usage is Bitcoin. Nowadays, with its development, a new platform called Ethereum becomes popular. Instead of supporting only cryptocurrency transactions in Bitcoin, Ethereum is a Turing-complete system which enables the end-developers to build software on "a hitherto unexplored compute paradigm"[3]. The platform provides a general-purpose Blockchain, and developers could customize their applications on it. As a result, the Ethereum platform is suitable for this project, which requires collecting and accessing the personal data other than the crypto-currency transaction records.

2.3 Encryption

Encryption method is used to encrypt data to some "random" information, so that the data is protected as it reveals no useful information before it's decrypted. There are mainly two classes of encryption, symmetric and asymmetric encryption. Symmetric encryption uses the same key for both encryption and decryption, while in asymmetric key scheme, different keys are required in encryption and decryption process, and the keys for encryption and decryption are uncorrelated. Usually, those keys for encryption and decryption are very big integers, e.g. 1024 or 2048 bits long, so the encrypted data will also be very large. While symmetric key system could be more efficient applying to large data, asymmetric key system is more secure, as it does not require key sharing between the two parties.
In this project, an encryption scheme with double trapdoor decryption mechanism[4]

will be implemented. It's an asymmetric key scheme, in which two different and uncorrelated decryption keys (available from two parties) are required to get the final information. Also, it allows some computations on the encrypted data, e.g. we could apply sum or average function, or get the top x answers without decryption.

3 Objective

- Implement the encryption and decryption scheme.
- Implement the big integer calculation on the Blockchain, to enable miners to analyze the encrypted information.
- Design the protocol for data verification and update with multiple miners.
- Design the scheme of the whole process, including data encryption, data collection and data analysis through Blockchain.
- Build the system and test the scheme.

A final deliverable will show the scheme and its performance analysis.

4 Methodology

This section discusses how the project will be implemented and the technologies used.

4.1 Ethereum

We will be using Ethereum as the Blockchain platform in this project. As discussed in section 2.1, Ethereum supports development of decentralized applications by developers. Like other Blockchains, Ethereum provides a cryptocurrency called Ether and supports some transactions on the cryptocurrency. On top of that, Ethereum is also programmable, on which programmers could develop their applications and release the codes to the Blockchain, and users could run those customized codes to achieve some specific transactions. This feature of Etheruem makes it possible to collect and update any kind of data on Blockchain, other than just the cryptocurrency transactions records.

The programming language to develop Blockchain applications on Etheruem is called Solidity. Developers write Solidity codes and wrap the related variables and functions together in a smart contract. A smart contract could be regarded as Ethereum's storage and execution unit of codes. Users of Blockchain execute functions on smart contract to achieve some functionalities.

Another issue is the testing of the platform. It's not realistic to test the codes on the real Blockchain, on which each transaction will be charged for verification. As Ethereum has already made the source code open to public, we could build a private Blockchain on the local machine using its source code. For small scale project, the online editor of Smart Contract implemented by Ethereum platform could be useful, which provides a virtual machine to build a private Blockchain.

4.2 Verifiable Blockchain

As mentioned in section 2.1, a very important issue of the project is to make sure the honesty of the miners, who are responsible for collecting and analyzing the updated information through a period of time. If the miners provide fake information, all users of the Blockchain are cheated, as they update whatever the miners broadcast to their local database. Verifiable Blockchain is a solution to the issue, which could ensure the authentication of the information. Currently there are researches on verifiable Blockchain, and some solutions for some specific information updating methods are available. We will be doing research on this topic, and provide a verification scheme for this project.

4.3 Data Collection and Analysis

For data collection, the data provider will firstly encrypt their data using the public key on their local machine, then provide the encrypted data to the Blockchain, which is public to all Blockchain users. This will ensure the privacy of the data. After miners calculating the related statistics (such as the sum, average or top x of the results), the data requester will get those statistics from Blockchain and decrypt the information using their private key to obtain the final information. The encryption and decryption scheme are proposed to be implemented using Python, which supports big integer calculation.

For data analysis, as the encrypted data are very large, e.g. 1024 or 2048 bits long, and the Ethereum Blockchain currently does not support such big integer calculation, we will firstly implement a big integer calculation library on the Blockchain. Afterwards, miners will be responsible for calculating the related statistics using this library, and some verification scheme will be applied to help update the statistics.

5 Schedule and Milestones

Sep, 2019	• Research on the related topics and get the whole idea of the project by communicating with supervisor. • Finish the project plan. • Design the website of the project.
Oct, 2019	• Research on the verifiable Blockchain. • Write smart contract for collecting data using Blockchain. • Implement the encryption and decryption scheme on Python. • Implement the big integer calculation library on Blockchain.
Nov-Dec, 2019	• Collect and process data. • Implement the miner's function on calculating the statistics. • Test the whole process of collecting data, analyzing data and accessing data with one miner.
Jan-Feb, 2020	• Research on the verifiable Blockchain and come up with some plan for multiple miner verification scheme. • Prepare the interim report and interim presentation. • Update the project website.
Mar-Apr, 2020	• implement the verification scheme and finish the whole platform for this project. • Test the protocol and platform and get its performance. • Prepare the final presentation and final report.
May, 2020	• Update the project website. • Prepare for the poster session if required.

References

[1] C. Michael, Nachiappan, P. Pradan, V. Sanjeev, and K. Vignesh, “Blockchain technology: Beyond bitcoin,” *Applied Innovation Review*, June 2016.

[2] I. Marco and L. Karim R., “The truth about blockchain,” Jan-Feb 2017.

[3] W. Gavin, “Ethereum: A secure decentralised generalised transaction ledger,” *Ethereum project yellow paper*, April 2017.

[4] E. Bresson, D. Catalano, and D. Pointcheval, “A simple public-key cryptosystem with a double trapdoor decryption mechanism and its applications,” in *Advances in Cryptology - ASIACRYPT 2003* (C.-S. Laih, ed.), (Berlin, Heidelberg), pp. 37–54, Springer Berlin Heidelberg, 2003.