Private Data Storage on the Blockchain
15-300, Fall 2017

George Lu

November 3, 2017

1 Project Description

For my 15400 project, I will be working with professor Vipul Goyal on blockchains. My project will be building a secure storage system on the blockchain with various forms of access control. So the motivation behind this is the growing popularity of blockchains due to their use in cryptocurrencies, and some of the features they provide. Most notably, the immutability of blockchains and their level of fault tolerance, along with simply being a distributed ledger mean that they provide a unique form of record keeping. Unfortunately, as it stands, current blockchain implementations is that all of the data stored on it is public, which limits their use in cases where we may want private but still distributed data.

For the project itself, there are a few different access control schemes which are worth exploring. Perhaps the simplest to implement based on our current ideas is a time-based access control scheme, in which some private data on the blockchain becomes public at some fixed date/time. Very similar to this is data publication after some real-world event, say if a judge orders it so. Further extensions include access based on credentials/keys, and even further along, data access dictated by arbitrary Turing-complete programs.

So the current envisioned approach to this problem is modifying currently existing secret-sharing schemes (e.g. Shamir Secret Sharing) to accommodate the special necessities of the blockchain environment, then distribute data to miners for them to ‘publish’. So some of the challenges of this project are that we do not want miners to have any more access to the data than the general public, and that unlike in traditional secret sharing, miners here may have varying computational capabilities which create an asymmetry in the amount of data they can handle. In addition, some fraction of miners (assumedly less than half) can be actively corrupt, and publish incorrect information. Finally, we want to consider the usability of such a system. This includes keeping data storage and publication reasonably efficient, as well as considering incentivizing miner participation, perhaps in the form of integrating the data storage system with a cryptocurrency.

As well as what was mentioned previously, secure data storage on the blockchain could be the first step to realizing private computation on the blockchain. This is an extension of existing technologies like smart contracts except where the computation itself is kept private. In addition to the clear increase in flexibility this provides over the already increasingly popular smart contract system used in Ethereum, it also has some interesting theoretical implications.
Most notably, we can replace many cryptographic assumptions requiring a trusted third party to instead have an honest, but untrusted majority, which can lead to better non-interactive zero knowledge, program obfuscation, etc.

2 Project Web Page

http://www.andrew.cmu.edu/user/gclu/

3 Project Goals

75%: Develop and implement time-based access control, including the secret-sharing, error-correction, and share-division.

100%: Complete previous goal, and explore other forms of access control and implement them as well. Have an analysis of the security of the system, and look into auxiliary improvements such as the incentive system and improving performance

125%: In addition to the previous goals, explore access control based on arbitrary Turing-complete programs. Possibly begin research into its applications in private computation on the blockchain.

4 Milestones

4.1 15-300 End of Semester

By this point, I hope to find and familiarize myself with a blockchain implementation to the point I am comfortable working with it. I will also do further reading into the techniques that we plan on using, and have a solid understanding of the algorithms and proofs involved.

4.2 15-400 Bi-Weekly

January 31st: Data Encryption / Decryption in integrated to blockchain implementation
February 14th: Implement secret sharing schemes and error correction in program
February 28th: Have share division worked out and modify secret sharing to be weighted
March 21st: Prove security of access control scheme implemented thus far
April 4th: Complete implementation of time-based access control and develop incentive scheme for miners
April 18th: Add credential based/real world event based system and look into performance
May 2nd: Investigate program-based access control

5 Literature Search

I’ve looked at a writeup Professor Goyal has given me. In addition, I’ve read about Shamir Secret-Sharing and error correction codes as two major techniques which will almost certainly be used. I have also read about blockchains in general, and how they function, and some details on specific instances of blockchains such as bitcoin and ethereum.
6 Resources Needed

We will need a working blockchain implementation that we can adapt to our needs. We are currently still evaluating different options for what the best option might be, but this mostly involves examining different existing implementations on the web. In terms of hardware, nothing much should be needed for the majority of this project. The only case I am currently aware of where it may be helpful to have additional machines is for testing the performance of our blockchain in more large scale settings, in which case we may want many instances of the blockchain running simultaneously.