EE5003 Research Log

Masters in Electronic and Computer Technology 2020/2021

Student Name: Michael O’Hara
Student ID: 16414554
Project Title: Blockchain and its role in file storage

Please read before making entries in this log

Please use this Research Log to capture your continued research reading and its influence on your project design and implementation.

Be selective about what you record in this log. Do not use it as an informal notebook while you are reading a new paper. Only make an entry after you have read a paper that you consider important to the development of your project solution. It is expected that, by the end of the project, you will have made **between 10 and 20 entries (20 maximum)**.

Share your log with your supervisor for viewing throughout the project. You will submit the final version of the log for grading, at the end of the project implementation period. It will be assessed on the basis of

1. how well you have understood and presented your research problem;
2. how well you have used your analysis of the literature to inform your project design, implementation and the evaluation of your project results;
3. how well you have documented your scripts and development works in your git repository.

The Research Log contributes **15%** to the overall project mark.

Note: All entries you make in this log must use the prescribed format shown on the next page. You will maintain other notes as you progress through your project but they should not be recorded here. Fill in the details where the *** signs are.

Statement of project problem / research question (maximum 200 words)

This statement should be periodically reviewed and updated, as necessary, as your project progresses and you gain further insight into the detailed project challenges, requirements and objectives as your project work moves from background reading, literature review, initial project design planning and detailed design and implementation. Initially, start by stating your current understanding of the project objectives. After each meeting with your supervisor, review and refine your project problem statement, as required.

***

File storage is an ever-present requirement when using a computer system in today's day and age. One of the challenges to this requirement is the need for secure and valid protection from outside or malicious parties attempting to compromise the content of files stored on the system. The goal of this project is to create a blockchain network which can apply the beneficial facets of blockchain to the storage of contents of a file from a user's computer in a safe and reliable manner which, when compared to traditional storage methods can be deemed to be safer of the two options. The solution will be developed using the Hyperledger fabric network and the use of a Virtual Machine.

The link to your project's git repository

Please include all your scripts and the development works in your git repository. Please also make your git repository accessible to your supervisors and the module coordinator.

***

https://github.com/oharam29/EE5003_IoT_Project

Name of the first reference paper **(please add any following reference papers using the same template provided here)**
*****Implementation of Distributed File Storage and Access Framework using IPFS and Blockchain**
Summary of paper (maximum 100 words)
*** This paper is the study of a file storage solution built using a combination of IPFS and blockchain. The paper introduces the concept of IPFS and the reasoning for its use in such as system. It then goes on to discuss how it aims to use the combination of the two technologies to reduce storage sizes and add content-addressed based access of transactions
How is this paper relevant to solving your project problem or addressing your research question? (maximum 100 words)
***This paper was relevant in the solving of the project problem as it introduces the state of the art for the use of blockchain in the area of field of data storage. It also brings the to light the major drawback of using blockchain is the data size for transactions. This paper aims to reduce the storage of transaction size of a block in blockchain.
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
***The primary strength of this paper's solution is the use of IPFS. PFS is a distributed data storage which provides content addressed and allocates the unique hash for stored file. The hash created is only 46 bytes long which drastically reduces the amount of space needed to store the transaction and the information related to it. The weakness of this solution is that is designed only designed for images and videos thus the back-end code is not designed to handle text file and the like.

Name of the second reference paper (please add any following reference papers using the same template provided here)
*****Ensuring Data Integrity Using Blockchain Technology**
Summary of paper (maximum 100 words)
***This paper delves into the methods used by Blockchain to ensure that the data stored within the chain's integrity is valid at all times. However, the commonly used strategies that are employed to ensure this are quite bandwidth heavy and when combined with blockchain become unfeasible. This paper compares two alternate approaches, that being a third-party auditor or Provable Data Possession. The method used by the authors is an amalgam of the two to require hashes that are compared to evaluate the integrity of a transaction
How is this paper relevant to solving your project problem or addressing your research question? (maximum 100 words)
***For Blockchain to be used as a replacement for traditional storage methods the integrity of the system must remain valid at all times protect the data. This paper serves to show that the use of blockchain itself already provides a large defence to integrity as in order to crack the chain a quantum computer is required to break the cryptographic keys
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
***The strength of this solution is the actual data sent in the payload. Each type of transaction requires a minimum of 5 parameters to go along with it such as the user's signature, the actual and previous hash of the file. These can each be used to verify the integrity of a transaction and operating as if any do not match a known user, it can be assumed the transaction is an outside party. The weakness of the solution is the verification of the hash on each transaction adds extra time to the execution to transactions which will slow down operation of the network

Name of the third reference paper **(please add any following reference papers using the same template provided here)**
*****DECENTRALIZED SECURE CLOUD STORAGE USING BLOCKCHAIN**
Summary of paper (maximum 100 words)
***This paper aimed to create a system that can be used to store client data on chain. It is done so by storing parts of a single encrypted file, with each separate part being stored on a different node. The reason for this is it allows only the person who owns the data to recreate the file.
How is this paper relevant to solving your project problem or addressing your research question? (maximum 100 words)
***This paper was relevant to the solving of my project problem as the problem area of this paper was very similar to my own. It was focused on storing files for a user in a secure manner and allowing the user to retrieve them. This paper gave me a good insight into the methods that I should undertake to achieve this. It also underlined some good technologies for achieving my goals.
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
***The main strength of the solution outlined in this paper is that it is built on Ethereum blockchain technologies which offer a faster transaction time that most blockchain networks. The weakness of the solution is that there is no built-in redundancy for a node failure meaning that if a node storing part of the data fails there is no way to recover that portion of the data.

Name of the fourth reference paper **(please add any following reference papers using the same template provided here)**
*****An improved P2P File System Scheme based on IPFS and Blockchain**
Summary of paper (maximum 100 words)
***Similar to the first paper this paper is a system built using both Blockchain and IPFS. This paper also includes the use zigzag codes to ensure data reliability and availability. Zigzag codes are a traditional instance of erasure codes which divide the block in *k* blocks and are encoded into *n* blocks. The result of the paper was that individual users no longer suffer from high throughput issues and can store files on a chain reliably.
How is this paper relevant to solving your project problem or addressing your research question? (maximum 100 words)
***This paper was relevant as it highlighted once again the need to find a way to reduce the amount of data needed for a transaction in order for a network to function at peak efficiency for the longest amount of time to cater for the maximum level of usage
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
*** The strength of the proposed solution is that the improvements made to the IPFS architecture will benefit the users and provide relief from the high throughout issues and provide benefits to the content service providers. It also allows users to interact with the system with ease without the need to maintain a node at full function efficiency.

Name of the fifth reference paper (please add any following reference papers using the same template provided here)
*****Metadisk: Blockchain-Based Decentralized File Storage Application**
Summary of paper (maximum 100 words)
***The goal of this paper is to prove that cloud storage applications can be made more decentralized, more secure, and more efficient through the use of the open-source software project Metadisk. Metadisk serves as the non-technical user interface as well as development platform for the Storj network. Using the Metadisk web interface or API, users may securely upload and download their files from the network and th network also has built in redundancy to prevent loss of data.
How is this paper relevant to solving your project problem or addressing your research question? (maximum 100 words)
***This paper was very relevant to the solving the project problem as it uses hashing to identify the files stored on the network and the files are encrypted before reaching the network it solves the weakness to the man in the middle attack. This was the inspiration for the methods used in relation to hashing the file before upload.
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
***The weakness of the solution proposed is that at the time of publishing the system is bottlenecked by a “Satoshi” style blockchain to only allowing 7 transactions per second. This would cause an issue if applied to a medium to large scale business needing to handle a lot of concurrent requests. The solutions strength is that the time to complete a transaction is relatively fast which can be used to combat the bottle neck mentioned previously.

Name of the sixth reference paper **(please add any following reference papers using the same template provided here)**
*****Secure Data Storage and Recovery in Industrial Blockchain Network Environments**
Summary of paper (maximum 100 words)
***This paper is about the implementation of blockchain for a fault-tolerant distributed storage and recovery algorithm. It goes on to explain the causes for node failures and the consequences of said failure. Then the algorithm used to combat this is explained and results show that it is capable o completing what it was designed for.
How is this paper relevant to solving your project problem or addressing your research question? (maximum 100 words)
***This paper highlights the need for methods for data repair and recovery in the event of a node failure a redundancy is necessary to prevent the loss of data stored within the node. It also helped in the decision of blockchain technology to be used in the project as the Hyperledger Fabric network has a few built-in redundancies for failed nodes.
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
***This paper demonstrates how the use of regeneration that is coded in simple manner can provide a very high level of repair capability. This is vital for when a node such as the nodes in a blockchain network fails and there is need to recover the data stored on the node. This system accomplishes all the goals it sets out to do and all tests on the system show a better level of performance in aspects such as data storage and repair rate increasing by 9% and 8.6 % respectively

Name of the seventh reference paper (please add any following reference papers using the same template provided here)
*****A Study on Improvement of Blockchain Application to Overcome Vulnerability of IoT Multiplatform Security**
Summary of paper (maximum 100 words)
***This paper takes a deep look at the security weaknesses of existing Internet of Things devices and methods of overcoming these weaknesses using Blockchain technologies. The main way of doing this was by building Blockchain of Things ecosystem.
How is this paper relevant to solving your project problem or addressing your research question? (maximum 100 words)
***This paper was a good reference for the duration of the project as even though it was a lengthy paper it gives good insight into the vulnerabilities of IoT devices and gave a number of possible solutions to these weaknesses.
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
***The solution offered by this paper is an ID-based encryption authentication protocol rather than the standard PKI-based authentication. It is a strong solution to the weakness that arise from IoT as it is more secure than the existing PKI-based one, but it has been proven to be more light weight. From their testing this approach appears to be 185.7% faster at registering new devices and the overall operation speed is 83% faster. The application and introduction of a system like this would be very costly however meaning it may not be feasible to go widespread

Name of the eighth reference paper (please add any following reference papers using the same template provided here)
*****When Blockchain Meets Distributed File Systems: An Overview, Challenges, and Open Issues**
Summary of paper (maximum 100 words)
***This paper is a deep comparison of the traditional distributed file systems (DFS) an the newly emerging technology of file systems that integrate technologies like IPFS and blockchain.
How is this paper relevant to solving your project problem or addressing your research question? (maximum 100 words)
***This paper was one of the most important resources for the duration of the project as it offered a great insight into the open challenges of the use of blockchain for file storage such as the scalability performance, application issues and big data issues and allowed me to try to create my own solutions to these challenges.
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
***The paper highlights that the introduction that the combination of IPFS and Blockchain are able to alleviate the issues with data latency when interacting with big data due to the characteristics of secure data storage and the decentralised nature of the storage however it was discovered that downloading and resolving operations could be bottlenecks while IPFS clients are reading objects from remote nodes which would be costly in the long run delaying critical actions

Name of the ninth reference paper **(please add any following reference papers using the same template provided here)**
***** A Blockchain-based Decentralized Data Storage and Access Framework for PingER**
Summary of paper (maximum 100 words)
***The goal of this paper is to develop a system in which metadata of files is stored on the chain while the actual files are stored off chain through the use of Distributed Hash Table. The reasoning for this allows for decentralised storage and lookup capabilities.
How is this paper relevant to solving your project problem or addressing your research question? (maximum 100 words)
***This paper is useful to the project as it is built upon the Hyperledger Fabric network which is the same network being used in the project and the paper provides useful insight into the requirements for development on the network
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
*** As the network can be built on the Hyperledger fabric network it means there is no requirement for a crypto currency blockchain, and it offers a modular approach depending on the requirement of the system.

Name of the tenth reference paper **(please add any following reference papers using the same template provided here)**
*****On the Exploitation of Blockchain for Distributed File Storage**
Summary of paper (maximum 100 words)
***This paper designs a system comprised of 3 parts which are the nodes, users, and services. This system will allow the users to store their files across multiple nodes and encrypting them with their private keys. Once developed the authors then test the effectiveness of this system.
How is this paper relevant to solving your project problem or addressing your research question? (maximum 100 words)
***The paper is relevant to the project as the system developed is similar to that of the project. It uses a unique key to query the files on the chain, this technique was adopted for this project also to allow the user to find their own files.
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
*** The proposed system was vigorously tested by the authors and the test show that the data written, and the throughput are better than that of the previous system iterations. The result show, the improvements are quite significant, but these tests were carried out on a very powerful computer which has specification which could largely be determined to be well above that of an average user and thus the tests do not accurately show how a regular users interaction with the system would fare.

Name of resource used (please add any following reference papers using the same template provided here)
*****Hyperledger Fabric**
Summary of resource (maximum 100 words)
*** Hyperledger fabric is an open-source, non-cryptocurrency based distributed ledger software. It serves as the foundation for the development of applications with a modular architecture. This allows the network to deliver confidentiality, resiliency, flexibility, and scalability
How is this resource relevant to solving your project problem or addressing your research question? (maximum 100 words)
***This resource is the groundwork for the project. It is the distributed ledger that the files and metadata about them will be stored upon. It was chosen because of its non-crypto currency-based nature so there are no transaction fees and ease of use. The software is also modular in design meaning the chain code it runs can be seamlessly swapped in and out depending on the project meaning chain code for file storage can be written package and integrated to carry out project goals.
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
***The main drawback to this resource however was the steep learning curve going in with little knowledge beforehand, there are a lot of settings and permissions which must first be given on a machine in order to run the network. But the modularity offered fair surpasses this drawback in terms of importance when deciding on the usage of the resource.

Name of resource used **(please add any following reference papers using the same template provided here)**
*****Virtual Box VM**
Summary of resource (maximum 100 words)
*** General purpose full virtualizer for x86 hardware, targeted at server, desktop, and embedded use. This allows a user to run a user Linux based guest OS on a Windows based host machine.
How is this resource relevant to solving your project problem or addressing your research question? (maximum 100 words)
***This resource allows for the easy running and maintaining of the Hyperledger network. This is a cumbersome process on a windows machine. It also allows for the easy installation of the prerequisites such as Node.js and Golang.
What are the strengths and weaknesses of the solutions/methods/technologies proposed in this paper? (maximum 100 words)
***The strength of this resource is that if offers easy installation of prerequisites and the fact that the machines specifications can be adjusted to the needs of the project. For this project a large storage and memory capacity is required and the VM allows this to be achieved.