Feasibility Study of the Usage of Blockchain Technology in Online Privacy Protection

Ben Folk-Sullivan
Department of Computer Science, Earlham College, 47374
Richmond, Indiana
bdfolks14@earlham.edu

ABSTRACT

In the modern day, a lot of private information people want to protect can easily find its way onto the internet. Blockchain Technology is a computer design architecture created to keep information secure and immutable. To find a way to better protect private information, this study explored the feasibility of using a blockchain system to protect online privacy. This study focuses on regulating access to browser cookies by online web-servers. Before a web-server can access a cookie, a blockchain network will authenticate the web-server's permissions. Request Latency, Transaction Latency, and Transaction Throughput scale with blockchain members used to determine this set-up's feasibility. Due to continued technical difficulties related to the structure of the coding and available resources, the study could not be completed.

KEYWORDS

Blockchain, Privacy Protection, Undergrad

1 INTRODUCTION

Blockchain is a new technology developed for digital currency. Every member of a system work together to maintain a unified, publicly accessible ledger of their transactions. This ledger is made up of a chain of immutable entries, or blocks, are that are linked by cryptographic hash keys, and serves as a way to ensure the legitimacy of all transactions. No transactions through this system can be conducted with falsified resources by hostile parties.[3]

In recent years, the scale and complexity of the Internet of Things has increased by leaps and bounds. It has become an integral part of life in the modern world, but yet it has few protections. Users leave clear, detailed histories of themselves on both their own databases and those of dozens of websites and Internet Service Providers (ISPs). They have no control over this information or the profiles made from them, and the information has turned into a good traded and sold among companies to improve their marketing campaigns and revenue streams.[2] In a very real way, the average internet user is turning into a corporate product. Some countries are moving to implement legal protections for their citizens, such as the European Union's General Data Protection Regulations. However, such measures are unreliable and fallible. It is not hard for unscrupulous and hostile individuals to skirt the law and exploit weaknesses inherent in the system. But what can stop and stall them would be architectural structures meant to protect legitimate users while stymieing hostile users.

This study was conceived as a way to investigate one such structure, using blockchain technology to protect and regulate access to a user's personal information stored on their computer.

This study uses a series of servers all linked together in a blockchain network. One of these servers works as an internet website while the rest act as internet users. When a user goes to the website, it asks to personalize its display for the user, based off their cookie-file. The cookie file is stored on the user's system, and contains a host of data about the user's browsing habit, with the specifics varying from cookie to cookie. The website would access the cookie, and then used the values there-in to customize itself in a certain way. To access this cookie, the blockchain network will verify that the website currently has permission to access this file. Once permission is established, the website will pass several functions over to the user, which ran them on the cookie. The results are then passed back to the website. This structure also allows permission entities to tailor a user's experience based on their personal information, while keeping the information itself safely secured on the user's systems.

The rest of this paper is organized as follows. Section 2 gives a detailed overview of information needed to understand blockchain technology, cookies, and online privacy, as well as a summation of previous research done into these topics. Section 3 gives a system overview of the modules involved in the study and how they were implemented. Section 4 covers the experiments and the results summation. Section 5 is the summarization of the study, including the overall result and how research can go forwards. Section 6 is acknowledgements, followed by the References section.

2 PRELIMINARIES AND RELATED WORKS

Relevant to this study are blockchain technology, the internet cookie structure, and the state of online privacy. First in Section 2.1 this paper explains Blockchain technology and how it has been studied and used since its invention in 2008. In Section 2.2 gives a brief overview of the internet cookie and how it has changed over time. Section 2.3 talks about the privacy of user information online, it's current state, and some studies done to try and rectify the situation.

2.1 The Blockchain

Blockchain is a relatively new technology, with its uses only lightly explored outside of the cryptocurrency world.

Earlham College, October 2018, Richmond,IN, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Modern systems only keep track of transactions relevant to themselves, without looking into the wider context, which allows for parties to spoof their own assets, getting into places and obtaining data they shouldn't.Blockchain tries to rectify this. It is a list of 'blocks' which are linked together by crytographic hashes into a 'chain.' Each block contains the time of its creation and transaction data. While the contents of each block can be viewed by any system, or 'node,' hooked into the network, they cannot be modified in any way, shape, or form. When a new transaction is made, all members of the network reference this block through its cryptographic hash key, to see if the transaction is possible. If it is, the new transaction data and the key it used are stored in a new block, which has a key-reference to the old block. Thus, the blockchain serves as an open ledger with immutable proof of its authenticity, helping to ensure protection for the general user. [4]

The blocks that make up this chain are not created out of nowhere, nor is their number strictly speaking finite. When a new transaction takes place between the blockchain's member nodes, every system runs a series of protocols to validate the transaction, checking through the chain to see if the resources being traded are legitimate. All member nodes then vote on whether to approve the transaction or not, which each one having a different weight to their vote. How these votes are weighted varies between blockchain structures. One such structure is a 'proof of work' protocol, where the amount of work a node put into its validation process effects its vote weight. Another structure is 'transaction stake' weighting, where a node's weight is determined by how relevant a transaction is to it. All these protocols have their pros and cons, and none of them are immune to failure.

However, blockchain has a much lower risk of 'false' transactions. With this system double buying, the process of buying two distinct items with the same money each time, is hard to do with bitcoin, due to blockchain's inherent structure. To be able to produce a false transaction in a blockchain, several key systems in the network would need to be compromised, leading to a '51%' attack where more than half of the total vote weight has been compromised.[7] This however is much harder to do than simply compromising one system and having it produce false transactions, which is something that can happen in many modern structures.

Blockchain technology has been heavily explored in the financial world, and much of modern understanding comes from this field, and cryptocurrency has taken off as a result. The pros it offers are most immediately apparent there, and provide valuable insight into the the technology's inner structure.[6]

In the medical field, several studies have been done.[1] [8] In these studies, patient data was kept in private systems that could be accessed via blockchain keys, with these keys only being in the possession of authorized users, such as the patient's currently assigned doctor and nurse. In addition, other users of this system could access large swaths of patient data without knowing any of the names or persons the data was attached to. If a doctor wanted to know how many people were suffering from a specific condition displayed a certain symptom, they could get that data easily. This allowed for patient data to be easily kept and accessed to improve their care without risking confidentiality. Medical research was improved as well, because professional researchers could get the data they needed without intruding into patient files. Personal online information is valuable in much the same way as personal medical data. It contains extremely private and personally valuable information that can also prove important to large scale data studies.

2.2 Cookies

Cookie are the term for files of data generated by websites in relation to a user. This file is then stored on the user's personal hard-drive, and is accessed by websites with the proper protocols through the web-browser whenever the user visits them. The kind of data stored in a cookie varies, ranging from website settings to names, home addresses, and credit card information.

Originally, cookies were unique per website: even if a user had a cookie for siteA on their computer, siteB couldn't access it. However, this structure was detrimental to companies who placed advertisements on websites, as the data collected in siteA and siteB's cookie could not be used together. So, third-party cookies, tailored to the advertiser's code, were developed. These cookies work across sites, allowing an advertiser to share their information on a user across multiple sites, which meant they could better target their adds. This also allowed them to track a user's browsing habits, forming a more coherent picture of them. If a hostile party were to compromise one of these cookie providers, they could then view and gather this data on their own, gathering large amounts of information that users do not want them to have.

Ostensibly, a user does not need to store cookies when accessing a website, and can thus prevent these files from being created and stored on their computer, but in practice this doesn't hold true. Many websites lock content behind log-in protocols, or inform users that by using their site they automatically consent to having cookies stored on their browser. There is little reason for them to give-up the massive benefit that is the ability to better match a user's preferences.

2.3 State of Online Privacy

When on the internet, people leave a very clear trail of who they are, what they've done, where they've been, and so on. Many online entities want this information so they can better tailor their marketing and advertising. This has led to groups whose entire business model is based around gathering and collating information into bundles that they then sell for a tidy profit. This is a system that moves vast quantities of data between large corporations, very often over unsecured networks, and then stored with minimal encryption and security. Hackers, scam artists, identity thieves, and other hostile individuals, can access and collect this information with relatively little effort. From simply obtaining a credit card number to completely stealing all information related to verifying a person's identity, they can cause catastrophic damage will little fear of repercussion.

However, many non-hostile entities benefit greatly from being able to use this information, and have an active interest in staying able to use that information. Personal information should just be put out there for anyone to use, but there are a lot of benefits of giving it to trusted and reliable entities.

When it comes to online data protection however, less practical work has been done. The Internet of Things is constantly shifting, and what information is harmless and what is sensitive changes

as well. The fact that an individual spends a lot of time at a coffee shop, writing work reports and browsing fashion articles might not seem important, but the clothing chain that operates just down the street from that coffee shop would like to know what clothes the individual is interested in. ISPs, product companies, and more are incentivized to hunt down and gather as much data as they can on people.[9] On the macro scale this data is no different than other broad surveys taken of people's likes and dislikes, and can be a major contributor to social studies and research. But this data is associated with individuals, alongside things like their passwords, schedules, associates, and so much more.

A study was done with the use of a blockchain system model to protect online information, similar to this study.[10] They had a handful of mobile phones that used blockchain technology to store and access their browsing information on offchain servers, and could grant keys that would allow internet services to view what information the user's had made available for viewing. This research was limited in its scope and ability to actually protect information. They looked into using their system on a series of mobile phones, and focused on simply whether someone could or could not access a piece of data. Once permission to access the data was granted, nothing was done to stop someone from recording the information. In addition, their system did not keep a detailed history of requests, such as who requested to see what, if the request was granted or denied, etc.

3 DESIGN

To study the feasibility of blockchain technology to protect online privacy is, a system using a blockchain to regulate access to a cookie file containing user information was designed, as shown in Figure 1.

3.1 Overview

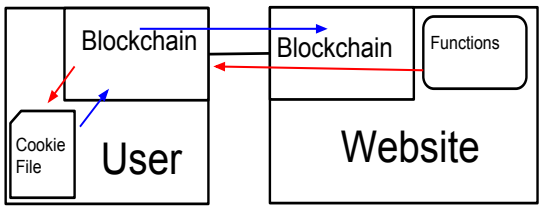

Figure 1: A diagram of the system framework

A single node in the blockchain network is labeled as a 'website' while all other nodes in the chain are labeled as users. The website requests access to a cookie on a user's system, trading over a key to signify it has permission to do so. The blockchain network will evaluate this key, checking to see if it is legitimate. If it is legitimate, the website will then pass over the function it intended to run on the cookie. The user's systems, which are themselves not connected to the blockchain, will then run those functions on the cookie, and pass the result back to the website. The website then loads the webpage, customized for the user based on the output. When the website first asks a user for permission to access its cookie, and the user accepts, the user will generate a blockchain item pair. This item is then transfered to the website. When the website makes subsequent requests of the user, the blockchain system will check to see it has the appropriate key and that it matches its pair on the user's database. These two items cannot be transfered away from either the user or the website.

The setup of this structure allows for all parties to benefit. Websites will still be able to customize themselves to their users, and advertisers will still have access to browsing data across multiple platforms, and can thus better target their interests. The personal information of users will be protected and guarded from hostile actors, as the data itself becomes blackboxed. The only things seen by other individuals is the output of their functions.

3.2 Implementation

The blockchain program used for the experiments is the open source program Multichain.[5] It was chosen for its platform versatility, being able to run on Linux, Windows, and Mac systems. The original plan was to run three sets of experiments with two, four, and eight member nodes in the network, hosted on the Earlham Cluster Servers. This became nonviable, as the way the servers were networked together caused the Multichain program to consider them, as a whole, a singular server, and thus would not let itself be run multiple times.

Instead, experiments are split into three groups. Each experiment in a group has the same number of nodes, and one node in every experiment would be labeled as the 'user,' with the others as websites. Websites granted permission to view the user's cookies are called 'yes-websites' while those not granted permission are called 'no-websites.' Each experiment in a group has a different assortment of yes-websites and no-websites.

There are three values to be measured in each set of experiments: Request Latency, Request Response, and Yes/No Ratio.

(1) Request Latency: the amount of time that passes from the website requesting to access the user's cookie to being granted or denied this response.
(2) Request Response: how many websites are requesting to see a singular user's cookie file at once.
(3) Yes/No Ratio: How many of the servers are 'Yes-Websites' and how many are 'No-Websites'

How these factors scale with different network sizes, and how they change with different ratios of 'yes-websites' to 'no-websites' is what determines the feasibility of this structure. With how the internet is ingrained into the modern day, it would be unreasonable to use such a structure on a large scale if there were significant issues related to time on even the small scale.

4 RESULTS

Due to technical difficulties and poor planning by the researcher, this study was unable to reach the testing stage. Therefore, this section is incomplete, but still provides an accurate summation of the planned experiments and what they could have meant. As mentioned in Section 3.2, the program used for the experiments is the open source blockchain program multichain.

4.1 Experiments

Group 1	Nodes	Yes-website	No-website
exp1	2	1	0
exp2	2	0	1
Group 2			
exp1	4	3	0
exp2	4	2	1
Group 3			
exp1	8	7	0
exp2	8	5	2
exp3	8	4	3

Figure 2: Experiment Tables

The experiments for the study are divided into three groups, based on the number of nodes for the experiments in that group, as seen in Figure 2. Keep in mind that each experiment has 1 node acting as the user.

Group 1 has 2 nodes, Group 2 has four nodes, and Group 3 has eight nodes. This provides a decent measure of scope in the limited testing environment, and the individual experiments within each group allows observation to see if the overhead is meaningfully effected by no-websites to yes-websites. The exp1 experiment of each Group tests the structure in an ideal environment, that all requesting websites have access to view the user's data. With the successive experiments, the ratio shifts in favor of 'no-websites' that do not have permission to view the user's data. This gives some idea of how the system framework performs in different levels of non-per missioned users.

In each experiment, every website makes an approval request of the user, and the Request Latency, the time between when the request is made and when it is answered, is then recorded into a results table.

4.2 Evaluation

The results obtained from running the experiments as detailed in Figure 2 are compared against the performance statistics of modern network structures. For each Group, an Group Request Latency would be calculated by taking a simple mean average of each experiment's Request Latency. If a Group's Average Request Latency is longer than a modern network of the same member-size, it is labeled as 'slower.' If the Average Request Latency is shorter, it is labeled as 'faster.' If the Average Request Latency is almost the exact same as a modern network, within a deviation of 1 second, it would be labeled as 'equal.'

The end result is a table of values that shows whether the proposed framework is significantly slower than the modern one for a network of specific sizes. If it is not, than the added security from the blockchain technology means the concept is feasible and sound.

5 CONCLUSION

This research is a first look into how feasible it would be to protect online privacy through the use of blockchain technology, and is sadly incomplete.

This study is incomplete, but there are several ways it can be taken in the future. The experiments and ideation were made with an eye towards the small scale and practical. In the future, with more resources, more experience, and more time, the concept could be expanded to work on a larger scale to better test the feasibility of blockchain technology in online privacy protection. These studies could be done with multiple different blockchain structures, seeing what impact it would have on Request Latency and Request Response.

Another direction future research can be taken is to explore failure management. The security of blockchain technology is already well tested and explored. However, how the system can handle and recover from security failures, which this system is not immune to, is an important area of study. What safeguards to have in case of failure would be practical, and is there unanticipated fallout from this failure that makes the structure itself worthless?

6 ACKNOWLEDGEMENTS

Thank you to my computer science teachers, Xunfei Xiang, Ajit Pai, David Barbella, and Charlie Peck for everything the taught me. Specific thanks to the first two for the help they provided me with during this research project. Thank you to Earlham College for 4.5 years of memories that have shaped me. Thank you to my parents Ann Folk and Luke Sullivan and my sister Katherine Folk-Sullivan for all the support and encouragement you've given me.

REFERENCES

[1] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman. 2016. MedRec: Using Blockchain for Medical Data Access and Permission Management. In *2016 2nd International Conference on Open and Big Data (OBD)*. 25–30. https://doi.org/10.1109/OBD.2016.11

[2] Elisa Bertino, Kim-Kwang Raymond Choo, Dimitrios Georgakopolous, and Surya Nepal. 2016. Internet of Things (IoT): Smart and Secure Service Delivery. *ACM Trans. Internet Technol.* 16, 4, Article 22 (Dec. 2016), 7 pages. https://doi.org/10.1145/3013520

[3] Weige Wu David Shrier and Alex Pentland. [n. d.]. Blockchain & Infrastructure (Identity, Data Security). (May [n. d.]).

[4] Joshua A.T. Fairfield. [n. d.]. Smart Contracts, Bitcoin Bots, and Consumer Protection. ([n. d.]).

[5] Gideon Greenspan. 2015. multichain program version 1.0.6. https://www.multichain.com/.

[6] Michael Mainelli and Mike Smith. [n. d.]. Sharing Ledgers for Sharing Economies: An Exploration of Mutual Distributed Ledgers (Aka Blockchain Technology). *Journal of Financial Perspectives* (November [n. d.]).

[7] Jennifer J. Xu. 2016. Are blockchains immune to all malicious attacks? *Financial Innovation* 2, 1 (10 Dec 2016), 25. https://doi.org/10.1186/s40854-016-0046-5

[8] Peng Zhang, Jules White, Douglas C. Schmidt, Gunther Lenz, and S. Trent Rosenbloom. 2018. FHIRChain: Applying Blockchain to Securely and Scalably Share Clinical Data. In *Computational and structural biotechnology journal.*

[9] Jan Henrik Ziegeldorf, Óscar García-Morchón, and Klaus Wehrle. 2014. Privacy in the Internet of Things: Threats and Challenges. *Security and Communication Networks* 7 (2014), 2728–2742.

[10] G. Zyskind, O. Nathan, and A. Pentland. 2015. Decentralizing Privacy: Using Blockchain to Protect Personal Data. In *2015 IEEE Security and Privacy Workshops*. 180–184. https://doi.org/10.1109/SPW.2015.27