AN EVALUATION OF SECURITY IN BLOCKCHAIN-BASED SHARING OF STUDENT RECORDS IN HIGHER EDUCATION

Timothy Arndt¹, Angela Guercio² and Younghun Chae²

¹Department of Information Systems, Cleveland State University, Cleveland, OH, USA
²Department of Computer Science, Kent State University at Stark, North Canton, OH, USA

ABSTRACT

Blockchain has recently taken off as a disruptive technology, from its initial use in cryptocurrencies to wider applications in areas such as property registration and insurance due to its characteristic as a distributed ledger which can remove the need for a trusted third party to facilitate transactions. This spread of the technology to new application areas has been driven by the development of smart contracts – blockchain-based protocols which can automatically enforce a contract by executing code based on the logic expressed in the contract. One exciting area for blockchain is higher education. Students in higher education are ever more mobile, and in an ever more agile world, the friction and delays caused by multiple levels of administration in higher education can cause many anxieties and hardships for students as well as potential employers who need to examine and evaluate student credentials. Distance learning as a primary platform for higher education promises to open up higher education to a wider range of learners than ever before. Blockchain-based storage of academic credentials is being widely studied due to the advantages it can bring. As with any network-based system, blockchain comes with a number of security and privacy concerns. Blockchain needs to meet several security-related requirements to be widely accepted: decentralization; confidentiality; integrity; transparency; and immutability. Researchers have been busy devising schemes to ensure that such requirements can be met in blockchain-based systems. Several types of blockchain-specific attacks have been identified: 51% attacks; malicious contracts; spam attacks; mining pools; targeted DDoS attacks; and others. Real-world attacks on blockchain-based systems have been seen on cryptocurrency sites. In this paper, we will evaluate the specific privacy and security concerns for blockchain-based systems used for academic credentials as well as suggested solutions. We also examine the issues for academic credentials which are stored “off-chain” in such systems (as is often the case). In this case, a Distributed File System (DFS) implemented with a peer-to-peer (P2P) architecture is often the choice for the storage of the academic credentials since it matches the decentralized nature of blockchain. Blockchain then contributes much to the usefulness of such a DFS, making it in turn a good match for a P2P DFS such as IPFS.

KEYWORDS

Blockchain, Mobile Education, Higher Education, Privacy, Security

1. INTRODUCTION

The blockchain idea was introduced in a paper published by Satoshi Nakamoto [1] and deployed in the bitcoin cryptocurrency the following year. Blockchain is an open, distributed ledger that can efficiently record transactions between two parties in a verifiable and immutable (permanent) fashion without the need for a trusted third party (disintermediation).
Bitcoin employs a peer-to-peer architecture and relies on proof-of-work, a piece of data time consuming and computationally complex to produce, but which is easy for others to verify (via “miners” who are rewarded with bitcoin for this computational work) and which satisfies certain requirements as a consensus mechanism. Consensus mechanisms allow for the correctness or “truth” of a transaction to be confirmed (depending on a set of rules) when multiple distributed actors perform transactions, and some of those actors may be untrustworthy. Subsequent developments have allowed blockchain to be programmed (via smart contracts) to trigger transactions automatically [2]. Transactions can cause code implementing rules which are stored on the blockchain to be run, through these smart contracts blockchain can lead to what have been called Distributed Autonomous Organisations (DAOs).

As a foundational technology, blockchain has been used or proposed in many application areas besides cryptocurrencies [3] including the banking sector [4], land registration (especially in developing countries) [5], the insurance sector [6], and electronic voting [7]. Alternative consensus mechanisms (such as proof of stake and mechanisms based on Byzantine Fault Tolerance) have been developed as well as alternative architectures (e.g., client-server).

Higher education academic records are typically stored and managed by academic institutions in a centralized manner. This leads to problems when students are highly mobile, attending several institutions. They must collect credentials from each institution. Also, when students transfer from one institution to another, their credentials must be evaluated manually. This is both slow and expensive. Blockchain has been suggested as a solution for these and other problems arising from centralized control of academic records in higher education.

In this paper, we will evaluate the specific privacy and security concerns for blockchain-based systems used for academic credentials as well as suggested solutions. For further study on blockchain in higher education, [8], [9] (related works section of this work), and [10] will be helpful. We also evaluate the issues for academic credentials which are stored “off-chain” in such systems (as is often the case).

The organization of the rest of the paper is as follows. In section 2, we give a brief overview of blockchain in higher education. Section 3 looks at blockchain security issues and how they affect the higher education domain. Section 4 examines the combination of blockchain and distributed file systems, which will likely be widely used in the domain being studied, while section 5 gives a concluding discussion.

2. Blockchain in Higher Education

In this section we will take a brief look at a few representative projects in the application of blockchain technology in the area of higher education.

Blockchain in higher education promises solutions to critical problems in the area. For example, the problem of fraudulent academic documentation. Blockchain should be able to be used to secure and verify academic documentation, without requiring a user to have access to an institution’s record keeping system [11]. This provides a quick and easy way to identify fraudulent academic documentation. Blockchain also enables automated mutual verification of academic records (transcripts, etc.) by academic institutions, avoiding the need for manual verification, by supporting the development of a standard representation of those academic records, outside the control of any centralized authority (due to the decentralized nature of blockchain). The importance of blockchain in higher education can be seen by noting that blockchain provides self-sovereignty, trust, transparency and provenance, immutability, disintermediation, and collaboration [12]. Student learning materials, such as exams and
programming assignments, can also be stored on the blockchain in a secure (encrypted and immutable) fashion.

A number of researchers have explored the use of blockchain to store university grades, i.e., university transcripts. A group at the University of Glasgow has developed a functional prototype for storage of student grades at the institution [13]. The platform chosen was Ethereum, hence it was built on a public blockchain. Based upon an exploratory, qualitative evaluation the authors found several tensions between the concepts of a university as an organization and of distributed autonomous organizations (DAOs) in Ethereum. Another project with a prototype implementation is [14], where a private BigChainDB blockchain is used for storage of student transcripts (not grades within a course as in the previously described research, though). Initial results were reported to be promising. Mahamatov et al.[15] describe a prototype implementation of a university transcript system using an Ethereum private blockchain and ERC-20 tokens (a standard for tokens, which are needed to carry out smart contracts, on Ethereum) on that blockchain. Students are able to read their grades, while professors and administrative personnel can record grades.

EduCTX [16] is an ambitious project for the development of a higher education credit platform based on the European Credit Transfer and Accumulation System (ECTS), a framework approved by the EU. The decentralized higher education credit and grading system can offer a globally unified viewpoint for students, higher education institutions, and other potential stakeholders such as prospective employers. A prototype implementation has been built on the ARK blockchain platform [17]. ARK is a public blockchain, but the authors transformed it into a private (permissioned) one by taking advantage of the flexible nature of ARK to change the parameters of the DPoS (delegate proof of stake) consensus algorithm used. Logic to ensure the validity of transactions on the blockchain has been defined. ECTX tokens represent credits that students gain for completing courses (analogous to the way that ERC-20 tokens are used in [15]). The authors will use the prototype system firstly at their home institution, the University of Maribor, and then at a select set of institutions of higher education. They anticipate that this or a similar system could potentially evolve into a unified, simplified, globally ubiquitous higher education credit and grading system.

A more theoretical investigation is given by [18] in which the architecture for the Disciplina platform for student records is described and an analysis of the main issues arising from storing student records in a blockchain is given. Their platform incorporates both private blockchains (maintained by individual institutions of higher learning) and public blockchains, managed by “Witnesses” who witness the fact that a private block was produced by a valid institution. A good, theoretical discussion of the problems of privacy, provability, and data disclosure in this context is given.

Besides traditional transcripts, blockchain is also being used or proposed for various alternative types of educational credentials. Blockchain enables permanent authentication and storage for a myriad of alternative credentials made up of diverse micro credentials, nano degrees, MOOCs, and certificates/badges from various types of training programmes. These credentials can then be directly controlled and managed by users [19].

Among the most well-known of higher education blockchain projects was that at MIT’s Media Lab which created blockcerts, a mobile app for educational credentialing built on Bitcoin [20]. At the Open University, researchers have developed the OpenLearn system built on the Ethereum public blockchain which awards OpenLearn badges for completing sections of a course and passing assessments [21]. The creators of that system have also developed a blockchain project to create a permanent distributed record of intellectual effort and associated reputational reward that
instantiates and democratises educational reputation beyond the academic community. Blockchain for Education [22] is another prototype system supporting the storage, retrieval, and verification of certificates via blockchain technology. Certificates are an important means of proving lifelong learning achievement in today’s environment, however they are susceptible to forgery. Blockchain helps to solve this problem. The prototype system uses Ethereum and its smart contracts to manage identities of registered certificate authorities and the hashes of certificates which are stored in a separate, centralized document management system, while the profile information of certificate authorities is stored using the Interplanetary File System (IPFS) distributed file system. Storing data off the blockchain allows it to be deleted as is required, for example, by the European General Data Protection Regulation (GDPR) for personal information. Other uses of blockchain in higher education have also been contemplated, including motivation, assessment, advising, etc. [23]. A prototype system for a blockchain based learning analytics platform built on Ethereum has been proposed as well [24].

3. SECURITY AND PRIVACY ISSUES IN BLOCKCHAIN

A number of different types of security problems have been noted in blockchain. In this section, we survey a few of these and how they relate to blockchain in education.

3.1. 51% Attacks

Blockchain relies on a distributed consensus mechanism such as proof of work (PoW) to establish trust. Unfortunately, the consensus mechanism is itself vulnerable to a 51% attack. This occurs when a single malicious miner or mining pools controls more than 50% of the total hashing power of the entire blockchain [25]. The malicious miner who controls this much hashing power has basically unlimited power to manipulate and modify the blockchain information including reversing transactions, changing the ordering of transactions and blocking transactions from being verified. A mining pool for the Bitcoin blockchain at one time reached 42% of hashing power before miners dropped out of the pool in order to stop a 51% attack from becoming possible [26]. In larger networks, achieving 51% of hashing power is quite difficult and expensive, but in smaller networks it is possible. For example, one of the Bitcoin forks, Bitcoin Gold, was the victim of such an attack in 2018 and then again in 2020.

51% attacks require a large investment in computing power and coordination in order to be carried out. While the effort might be worthwhile for the monetary gain associated with blockchains used for cryptocurrencies, for large public blockchains the effort will probably not be worthwhile for the advantages to be gained by manipulating educational credentials, unless some group wants to corrupt all of the credentials for an institution (for example), not just those of a single student. It is not to be discounted, though, that such corruption could occur as collateral damage from a 51% attack aimed at some other, more remunerative target.

3.2. Malicious Contracts

Ethereum is an open-source, public, blockchain-based distributed computing platform and operating system featuring smart contract (scripting) functionality [27]. Smart contracts are effectively programs that are stored in blocks on the blockchain. Predetermined conditions control when these programs are executed. A typical use is the automation of the execution of a contract involving multiple parties. All parties to the contract are immediately aware of the outcome of the contract without the need for an intermediary. Another use of smart contracts is the automation of a workflow where the next action in the workflow is triggered when certain conditions are met. Users of the Ethereum platform are ‘pseudo-anonymous’ and a single user
can have multiple accounts under multiple cryptographic identities. It has been shown [28] how ‘criminal smart contracts’ can facilitate leakage of confidential information, theft of cryptographic keys, and various real-world crimes. Atzei et al. [29], have presented a taxonomy of vulnerabilities of smart contracts in Ethereum including such vulnerabilities as stack overflow, exception disorder and unpredictable state. Various countermeasures to these malicious contracts have been proposed, including malicious contract detection using supervised learning [27].

For higher education blockchain, likely the most serious threat from this type of attack is the leaking of private credentials. However, it is also possible that a malicious contract exploiting the vulnerabilities identified above might falsely update a student’s credentials, corrupting them.

3.3. Sybil Attacks

In a Sybil attack, a single adversary controls many nodes in a system by creating numerous fake identities in order to gain a disproportionately large influence, thus subverting a reputation system[30]. The vulnerability of the reputation system to Sybil attacks depends on the ease with which identities can be generated, the openness of the reputation system to inputs from entities without a chain of trust linking them to a trusted authority, and finally whether the reputation system treats all entities identically. The adversary in a Sybil attack presents multiple identities to a P2P network in order to effectively control multiple nodes in the network, thus achieving a disproportionate level of control over the network. A defense against Sybil attacks can rely on validated identities issued by a trusted authority. The requirements for agents to present a trusted identity conflicts with the need for permission less open membership, though. Anonymity is also lost. Anonymity can be preserved by alternative defenses against Sybil attacks such as social trust graphs or by imposing economic costs as artificial barriers to entry. Proof of work, such as required in some blockchains is an example of this approach. An example of a blockchain development to defeat Sybil attacks is TrustChain [31] which includes a novel Sybil-resistant algorithm NetFlow to determine the trustworthiness of agents in an online community.

3.4. Eclipse Attacks

While a Sybil attack affects the whole network, in an Eclipse attack the attacker targets a particular user (or node). The attacker attempts to prevent the victim from having a true picture of network activity and the current ledger state. The attacker effectively duplicates the network for the attacker, presenting this fake network to the victim instead of the real network. This is made possible since in a decentralized network like blockchain, a node does not connect to all other nodes on the network simultaneously – instead it is limited to a small, select group of nodes, who in turn are in contact with others. The attacking node controls a botnet of host nodes and needs to ascertain the select set of nodes that the victim is currently neighbor with (able to directly communicate with). An Eclipse attack will allow the attacker to monopolize all of the victim’s connections – both incoming and outgoing [32]. This will isolate the victim from other users on the network. The attacker is able to filter the victim’s view of the blockchain or to cause the victim to waste computing power on obsolete views of the blockchain. The attacker is also able to use the victim’s computing power to conduct its own malicious acts. In the context of education blockchain, a student might prevent a potential employer from having a complete view of all of the student’s credentials.

4. BLOCKCHAIN AND DISTRIBUTED FILE SYSTEMS

One of the drawbacks of using blockchain to store academic records is the relatively high cost of storing data on the blockchain. In order to gain the previously discussed benefits of blockchain
while being able to store large amounts of data, researchers in various domains have recommended that in the use of blockchain, the large amount of data required by the domain be stored off-chain, while the blockchain itself is used to provide authentication mechanism for access to the data [34]. Due to the decentralized nature of blockchain, it is natural to match the blockchain with a similar file system – namely, Peer-to-Peer (P2P) distributed file systems (DFS) [35].

Given to the aforementioned high cost of storing data on the blockchain, it is likely that educational transcripts themselves will be stored off the blockchain, with only the credentials needed to access and verify them being stored on the blockchain. One popular mode of storing data off the blockchain is to use the InterPlanetary File System (IPFS). IPFS is a peer-to-peer distributed file system, so its architecture and management style are a good match for blockchain-based systems.

The IPFS is both a protocol and a P2P network for storage and sharing of data in a distributed file system. User-operators of IPFS hold a portion of the complete data, rather than having all of the data at a central repository. IPFS uses content-addressing. Each file is uniquely identified in a global namespace. A user can serve a file by content address and other users can find and retrieve that content from a user who has it (there will likely be multiple replicas of that content in the system) through the use of a distributed hash table.

Blockchain has been used together with IPFS in the Filecoin [36] system which is a cooperative digital storage and retrieval method. Blockchain makes up for some of the deficiencies of P2P DFS, namely, it supports the verification of publishers and the credibility of the downloaded content, and it provides an incentive mechanism so that the nodes taking part in the P2P network which store the files are rewarded for sharing their bandwidth and storage resources which lead to increased file availability and robustness. In a traditional web-based environment, Hypertext Transfer Protocol (HTTP) is used to access data. In the blockchain/IPFS combinations, the IPFS protocol with its secured hashing is used which provides security and privacy.

For examples of the combination of blockchain and IPFS, see [33], [34] for the use in storage of electronic medical records, and [40] for the use in the automotive insurance sector.

Other popular blockchain-based distributed file systems include Swarm [37], Storj [38], and PPIO [39].

In terms of security, the Sybil and Eclipse attacks previously mentioned can also apply to these P2P DFSs. Other specific attacks, such (D)DoS have also been noted.

5. CONCLUSIONS

Blockchain has been increasingly accepted in higher education as an alternative for the storage of academic records. A systematic literature review [10] shows that interest is continuing to grow, with the peak of interest occurring in the last complete year surveyed - 2020. The main advantage of blockchain in this context is that the records are not under the control of several institutions or of a centralized third party.

The current approach has each academic institution maintaining its own records, and students having to collect records from each institution they attended in order to present them to prospective employers. This system is both slow and costly (for students – institutions may make money by charging for the release of academic records). Blockchain solves the problem of speed (in addition to taking the records out of the hands of the institutions, as noted above). Some cost
is associated with blockchain, however, much of that cost can be borne by prospective employers who carry out transactions on the blockchain in order to have access to academic records. Another important application in this area where blockchain can help out is in the evaluation of academic credentials when students transfer from one academic institution to another. Currently, the academic records must be evaluated manually, a process which is both slow and expensive. With blockchain, the whole process can be automated, perhaps employing the use of smart contracts, making the process much more efficient.

An alternative approach could be to have a trusted third party hold all of the academic records, but then we still have the problem of trust. Not all countries have an institution worthy of trust, and even in those that do, not all students may trust them (no universally trusted institution). Blockchain does not require a trusted third party, so it solves this problem.

As was seen in the descriptions of the types of attacks in the previous section, many of the attack strategies require obtaining large amounts of computing power to be effective. Such an attack may be considered worthwhile in the cryptocurrency application of blockchain, but in the education sector, no such financial incentive is apparent for attacks on a single student’s credentials. Possibly the blackmailing of a large institution poses more of a threat, though given the higher payoffs available to criminals in other blockchain areas, that is still rather farfetched. The threats associated with malicious contracts seem to be more of concern in educational blockchain, since they may corrupt educational credentials.

It should be made clear that though this paper has examined various security issues related to the use of blockchain for the storage and exchange of higher education student academic credentials, such a system is still generally better than the centralized alternatives. Blockchain technology secures and authenticates transactions and data through the use of cryptography. The immutable nature of blockchain prevents the deletion of or tampering with the academic records.

In any case, it will be necessary for users of blockchain-based educational credential systems to keep up to date with the latest security issues in the blockchain area in order to guard against attacks, since those attacks, and the defences against them, are continuously evolving.

REFERENCES


AUTHORS

Timothy Arndt is Chair and Professor of the Department of Information Systems, Professor of Electrical Engineering and Computer Science, and Director of Business Analytics at Cleveland State University. He received a Ph.D. in Computer Science from the University of Pittsburgh. He is the editor of several journals and has been involved in the organization of several international conferences. His research interests include database systems, software engineering, e-learning, and blockchain.

Angela Guercio is Associate Professor in the Department of Computer Science at Kent State University. She received a Ph.D. in Computer Science from Kent State University, a M.S. in Computer and Information Sciences from the Knowledge Systems Institute, IL, and the Laurea degree in Computer Science “cum laude” from the University of Salerno, Italy. She has previously been an assistant professor at Hiram College, OH, adjunct research visiting professor at Lamar University, TX, and a senior research associate at the University of Salerno, Italy. She has been involved in the organization of several conferences and is the editor of several journals. Her research interests include e-learning, software engineering, multimedia languages, and blockchain.

Younghun Chae is an Assistant Professor of the Department of Computer Science at Kent State University and resides at the Stark campus. He received both his M.S. and Ph.D. from the University of Rhode Island. His research interests include trust management, data analysis, and cyber security. He has worked with the Digital Forensics and Cyber Security Center at the University of Rhode Island since 2010. He is now director of the Cyber security and Intelligence Analytics Lab at Kent State University at Stark.