Blockchain diplomas: Using smart contracts to secure academic credentials

Fabian Schär, Fabian Mösli

In this article we describe how blockchain technology can be used to secure academic credentials. We provide insights from a joint project between the University of Basel's Center for Innovative Finance and BlockFactory Ltd, and show some considerations that went into the concepts as well as an evaluation on how these blockchain diplomas perform in comparison with other diploma solutions.

1 Introduction

Fraud is a serious problem in academia. A simple Google search for «buy fake university degree» leads to millions of results. Some of the service providers promise to deliver fraudulent diplomas within 24 hours and prominently advertise the fact that they have been in business for over a decade. Anyone can easily buy forged credentials for as little as USD 100. Needless to say, these fake diplomas have the potential to significantly damage the credibility of institutions and higher education as a whole.

Diplomas are generally used as a signaling mechanism, i. e. to demonstrate the presence of certain abilities. This signaling mechanism becomes much less efficient when it gets harder and more costly to distinguish between legitimate and fraudulent credentials. It is safe to say that this is the case with physical diplomas. These documents are simply not secure. Even with additional security measures such as holograms or watermarks, it is possible to counterfeit these documents. What makes things even worse is the fact that most application documents are either being sent in electronic form or using photocopies that obfuscate the security measures and mostly render them useless. Anyone familiar with basic software tools can easily modify sensitive data on these documents, including grades and even the name of the graduate. As a result, potential employers have to check back with the university to verify if the diplomas they have received are real. This supposedly simple request kicks off a highly bureaucratic process and results in a significant administrative burden for the universities. Moreover, the person processing the request faces the challenge of providing reliable information while making sure to be compliant with data protection law.

Some universities try to tackle this problem by using a centralized database and an online form to automate requests. While this certainly is a step in the right direction and significantly improves the efficiency, there are still certain drawbacks to this

approach. In particular, a centralized database is a central point of failure. If someone succeeds in gaining access, this person could add new entries and edit or delete existing ones at his or her own discretion. Additionally, the diplomas' validity would be dependent on centralized infrastructure, meaning that diplomas could only be verified when the institution's database is available. If the database were to fail or if the institution were to disappear, the diplomas could no longer be verified and would therefore lose their usefulness. Outsourcing the responsibility by using a centralized database which is managed by a third party does not solve any of the afore-mentioned problems. On the contrary, it introduces additional dependencies and a severe lock-in. Ideally, there would be a shared database, managed by a large community, with everyone being able to autonomously verify the legitimacy of its records. That is exactly what blockchain technology can be used for and the starting point for our project (Grech & Camilleri, 2017; Jirgensons & Kapenieks, 2018).

This article is structured in four sections. After this short introduction (section 1) we proceed with section 2, where we provide an overview of the implementation. In section 3 we compare blockchain diplomas to other solutions and show where the advantages and disadvantages lie. In the fourth and final section we conclude.

2 Implementation

In early 2018 the University of Basel's Center for Innovative Finance partnered up with the Proxeus foundation and BlockFactory Ltd to secure course certificates on a public blockchain (Center for Innovative Finance (University of Basel) and Proxeus, 2018). We decided to use Ethereum[1] (Wood, 2014) Mainnet (the second largest public blockchain; Buterin et al., 2013). Ethereum allows small applications (smart contracts; Szabo, 1997) to be stored and executed in the blockchain network. In very simple terms, we add information on academic credentials to the Ethereum blockchain. Since it is a public database, anyone can easily look up the information and thereby verify if a specific diploma is part of the database. While using a permissioned ledger (e.g. run by a consortium of institutions) would also have been an option, using a public blockchain frees the participating institutions from the need to operate and maintain the respective infrastructure and services. Permissioned ledgers may offer better efficiency at the expense of superior immutability of a public blockchain – a trade-off we preferred to avoid (Zheng et al., 2017).

[1] There are other universities and organizations who have launched similar projects (University of Nicosia, 2019; MIT, 2019; Government Technology Agency (Singapore), 2019).

A simplified step-by-step model of the diploma registration process can be described in four steps:

(1) The university issues a pdf diploma. These diplomas look just like normal diplomas. Among other things, they contain the full name, student ID and the grade.

(2) The university computes a digital fingerprint *h* for each diploma file. The digital fingerprint is a 256-bit representation of the diploma.

(3) The university creates and relays a transaction containing *h*, with the goal to add *h* to the blockchain.

(4) The transaction becomes part of a valid block and is confirmed on the blockchain. As a result, *h* has become part of the blockchain.

The verification process works in a very similar way. Anyone who has received the diploma can recompute its digital fingerprint and compare the result to the values on the blockchain. If the value can be found on-chain and certain criteria are met, it serves as proof that the diploma is valid.

What is great about this process is that anyone can independently verify diplomas within a few seconds. There is no need for centralized infrastructure. However, the fact that anyone can join the network, look up values and even add new entries to the blockchain, raises some questions mainly about the documents' authenticity and data protection. The next few subsections describe how we have dealt with these issues.

2.1 Authenticity

In a public blockchain there is no central authority with special privileges or permissions. Anyone can join the network and issue new transactions. Transaction data eventually become part of the blockchain, meaning that anyone can add arbitrary information to the public database. It would therefore be a fallacy to assume that something must be true just because it is stored on a blockchain. In particular, an attacker could create a fraudulent diploma and add the digital fingerprint of this diploma to the blockchain.[2]

To ensure that only the university can manage its academic credentials, we use public key cryptography. To understand the concept, we need to provide some background information. Public key cryptography uses pairs of keys, that is a public key and a private key. Each individual (or institution) may autonomously pick a private key, i. e. a random number from an unimaginably large set, and use it to derive a correspond-

[2]For an introduction to blockchain see Berentsen & Schär, 2017; Berentsen & Schär, 2018.

ing public key (point on elliptic curve). While it is straightforward to derive the public key from a private key, it is computationally infeasible to invert the function and derive a private key from a public key. This is important since the private key serves a similar role as a password and must therefore remain secret at all times. The public key can be best compared with a user name. It serves as the individual's pseudonym in the network and can be disclosed freely. Thanks to the one-way characteristics of the derivation, the individual who has created the key pair remains in exclusive possession of the private key; despite having disclosed the public key.

The two keys have a mathematical relation that allows messages that have been encrypted with the private key to be decrypted with the corresponding public key (and vice versa). This property can be used as follows: Whenever an individual creates a transaction, this transaction message must be encrypted (signed) with the individual's private key. Since everyone is in possession of the corresponding public key, they will be able to decrypt the message with ease, and thereby receive proof, that the transaction has been issued by the individual behind the pseudonym. If the public key of the university is known, it can be easily verified if a specific diploma has been added by this university or by a third party.

2.2 Data protection

Academic credentials consist of personal data. We therefore have two conflicting goals. On the one hand, we want the information to be accessible and verifiable. When someone receives a diploma, this individual should be able to consult the blockchain and autonomously verify if the university has added these data to the public ledger. On the other hand, we do not want personal information to be publicly disclosed. Writing sensitive information, including full name, student id and grades on a public database, surely would not be a good idea nor would it be compliant with data protection law.

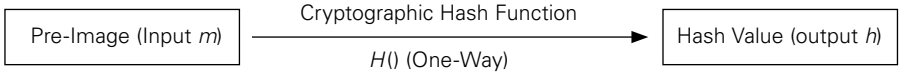
To circumvent this problem, we decided not to add any clear text information to the blockchain. Instead, we employ a special mathematical function and only store a cryptographic representation of the diploma on-chain, the digital fingerprint of the file. This so-called cryptographic hash function $H()$ is a deterministic one-way function that maps input data m of arbitrary length (pre-image), to a fixed-length output (hash value) $h = H(m)$. In our case, m corresponds to the diploma file and h therefore is the hash value of the diploma.[3]

[3] Computers represent everything as numbers. It is therefore possible to compute a hash value of text snippets or entire documents such as a diploma.

The term deterministic means that, given the same input, the hash function will always lead to the same hash value. In other words, there is no randomness involved. It is, however, important to point out that the results appear to be random, since even the slightest changes in the input lead to a completely different hash value. It is therefore not possible to deliberately generate hash values with certain characteristics, by picking the inputs.[4]

The term one-way means that it is infeasible to invert, i. e. it is straightforward to compute the hash value from the input but not the other way around. In other words, given *h*, it is infeasible to compute *m*. This relationship is shown in figure 1.

Figure 1: Cryptographic hash functions are one-way functions.



Now let us assume without the loss of generality that a university has issued one diploma. It uses a cryptographic hash function to compute the hash value of this diploma and then adds the hash value to the blockchain. Let us further assume that there are two types of observers: *A* and *B*. Observer *A* is in possession of the diploma and therefore is able to recompute the hash value and validate if it matches the hash value stored on-chain. Observer *B* is not in possession of the diploma. All he sees is the hash value that does not disclose any information on the original document.

The special properties of the hash function therefore allow anyone who is in possession of the diploma to verify if it is authentic. Those who are not in possession of the diploma will not gather any information from the hash value.

2.3 Smart contract

What we have described so far could be implemented without the need for a smart contract. As an example, one could use Bitcoin's Null Data (OP_RETURN) transaction type to achieve similar results. The hash values would be observable on-chain and,

[4]The function is non-injective, meaning that multiple elements of its input domain may be mapped to the same element of the output domain. In particular, since the input domain contains more elements than the output domain, we know that there must be collisions. However, due to the unbelievably large set of potential output values and the unpredictable effects when the input gets changed, it is infeasible to find any of these collisions.

thanks to the signature, one could also verify the source of the data. This very simplistic implementation would be sufficient, although somewhat cumbersome to work with. Moreover, this simple implementation would significantly limit our options.

For these reasons, we have decided to base our solution on a smart contract. A smart contract is a blockchain-based application that is governed by code and consists of a collection of state variables and functions. Its functions can be called by transactions. When called, they are executed in accordance with the contract's code and may change the state of the contract's state variables. We can use this contract to manage hash values of academic credentials.

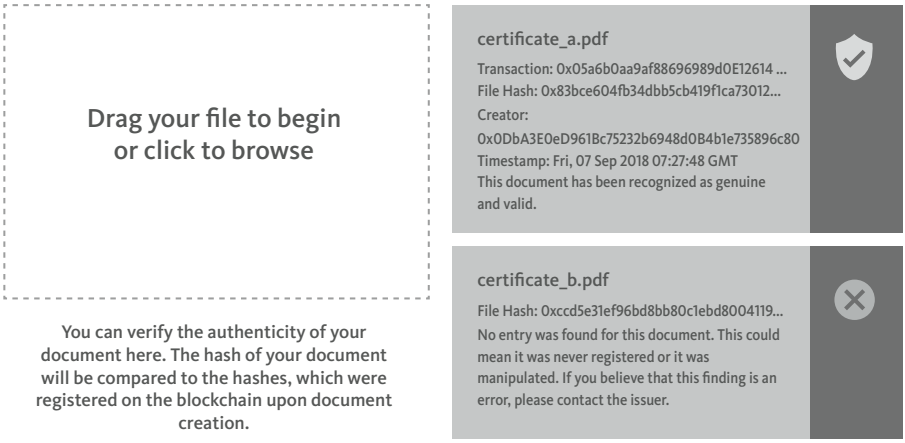
The main function of the contract allows new hash values to be added to the contract's storage. The contract accepts this addition only if the initiating transaction has been signed with the private key of a pseudonym that is associated with the university. Similar functions allow for the revocation of existing hash values and the assignment or removal of university representatives.

2.4 Verification tool

In theory, anyone can autonomously verify a diploma in his or her possession. The verifier has to install an Ethereum software client (e. g. geth) and thereby create a full node. He then has to download a copy of the entire blockchain and verify all transactions and blocks, compute the hash value of the document and compare the hash value to the ones stored in the smart contract. While it is great to have the option to verify the authenticity of academic credentials in a completely trustless environment and with no need to rely on anyone else, it is rather unlikely that, for example, an average human resource department will go through this process.

Well aware of this, we decided to provide a simple verification tool. The tool can be embedded in the university's website and is connected to an Ethereum node, such that anyone who receives a diploma can simply drag and drop the pdf on designated area to trigger a verification process. After a few seconds, the verification tool will show one of the messages shown in Figure 2.

Figure 2: Verification tool.



3 Comparison

Blockchain diplomas are just one among many solutions to secure academic credentials. Physical diplomas, certified electronic documents and similar projects implemented on a centralized database are some of the other options available. Accordingly, it is interesting to see how blockchain diplomas rank against these options and what advantages and disadvantages the solution has. The results are summarized in table 1 and discussed in the following subsections.

Table 1: High-level comparison between some of the available diploma options.

	Verification	Revocability	Timestamp	Cost Predict.
Physical Diploma	weak	no	none	fixed
Certified PDF	centralized	no	weak	fixed
Centralized Database	centralized	yes	weak	fixed
Blockchain Diploma	autonomous	yes	secure	variable

3.1 Verifiability

Academic credentials are only useful if they can be verified. We differentiate between autonomous and centralized verification. We call a verification process autonomous if someone who is in possession of the diploma can autonomously verify the authenticity of this document without having to rely on centralized infrastructure. We call a verification process centralized if the verification can only be conducted with the help of a central party (phone call, API call, web form or similar).

Blockchain diplomas are autonomously verifiable. The information is on a public blockchain, meaning that the process would work even if the issuing institution is no longer available.

The verification of diplomas based on centralized databases or certified PDFs rely heavily on centralized infrastructure. The same is true for physical diplomas. The process to verify these diplomas is usually not well designed, resulting in uncoordinated phone and email requests. Needless to say, these requests also depend on the existence of centralized infrastructure.

3.2 Revocation

Under certain circumstances institutions must be able to revoke academic credentials. If, for example, an irregularity is detected after a certificate has been issued, the issuer must be capable of declaring the document invalid. This can be easily achieved through an additional function in our smart contract that marks the respective diploma as revoked. Similar results are possible with a centralized database.

Physical diplomas struggle in this category. Once a physical diploma has been circulated, it must be assumed that numerous copies of this document exist. Without a database (blockchain or centralized) on which the status of the document can be updated, these copies are considered valid even if the original has been destroyed.

Certified PDFs can theoretically be revoked. However, in most cases it requires the issuer to invalidate the institution's certificate and therefore all academic credentials that have been issued with the same certificate.

3.3 Secure timestamping

A public blockchain can provide secure timestamps for any data. We can use these timestamps as an additional security measure to prevent backdating. If, for example, an institution's private key is leaked, a potential attacker would still not be able to issue diplomas from last year's class. We could add an additional criterion, i.e. define a specific time period during which the diploma has to be issued to be considered valid. If the diploma is issued during a different time period, it is considered invalid despite a potentially valid signature.

This is a big advantage over systems based on physical diplomas, which can be backdated indefinitely. In fact, it would be a rather straightforward task to recreate a diploma from 1975 with a different name. Even with certified PDFs and centralized databases one is at risk of backdating. If the certificate or the centralized database

gets compromised, a potential attacker may theoretically be able to perform any changes.

3.4 Cost predictability

The main disadvantage of our implementation is that the costs are not as predictable as with the other options. Transactions on the Ethereum Blockchain are subject to a fee. The fee is determined by the market, i. e. whenever there is a large queue of pending transactions, fees may rise. Failing to match current market prices may cause the transactions to get stuck for several hours or even days. While diploma issuance is usually not that time critical, this is certainly a category in which blockchain diplomas perform worse than some of the other options.

If the transaction fees rise to levels that are not sustainable for this application, we could combine several hash values in a so-called Merkle root or include all individual hash values of a semester in a master document and only add the hash value of the master document to the blockchain (University of Nicosia, 2019). Both approaches have the advantage that they would only require one blockchain transaction, at the cost that an individual would require additional information besides the diploma itself to be able to verify the diploma's authenticity.

There is some hope that Ethereum's scaling solutions (EthHub, 2019) will solve this problem before it even arises. However, we must be aware that the project is, at least to some extent, dependent on a timely arrival of these solutions.

4 Conclusion

We are content with the progress of the project so far and see blockchain diplomas as a valid option for issuers of academic credentials. The main arguments in favor of blockchain diplomas are the secure timestamping as well as the autonomous verification. The somewhat unpredictable costs are a disadvantage compared to other solutions, however, there are alternative blockchain-based implementations that would be significantly less affected by a large increase in Ethereum transaction fees.

Moreover, we believe that there is high potential for securing any sort of information on the blockchain. The same procedure could be used to prove the authenticity, integrity and content of any document.

Acknowledgements

Special thanks to Manuel Gall, Fabian Lindner, Katrin Schuler, Antoine Verdon, Artan Veliju, Silvio Rainoldi, David Matter und Lukas Karth.

References

Berentsen, A. and Schär, F. (2017), Bitcoin, Blockchain und Kryptoassets, BoD, Norderstedt.

Berentsen, A. and Schär, F. (2018), 'A short introduction to the world of cryptocurrencies', Federal Reserve Bank of St. Louis Review 100, 1 16.

Buterin, V. et al. (2013), 'Ethereum white paper. (2013)', URL https://github.com/ethereum/wiki/wiki/White-Paper .

Center for Innovative Finance (University of Basel) and Proxeus (2018), 'Certificates based on blockchain technology', https://cif.unibas.ch/fileadmin/user_upload/cif/press_release_EN.pdf.

EthHub (2019), 'Ethereum roadmap / ethereum 2.0 phases', https://docs.ethhub.io/ethereum-roadmap/ethereum-2.0/eth-2.0-phases/.

Government Technology Agency (Singapore) (2019), 'Opencerts', https://github.com/OpenCerts/opencerts-documentation/.

Grech, A. and Camilleri, A. F. (2017), 'Blockchain in education'.

Jirgensons, M. and Kapenieks, J. (2018), 'Blockchain and the future of digital learning credential assessment and management', Journal of Teacher Education for Sustainability 20(1), 145 156.

MIT (2019), 'Blockcerts', https://www.blockcerts.org/about.html.

Szabo, N. (1997), 'The idea of smart contracts', http://w-uh.com/download/WECSmartContracts.pdf.

University of Nicosia (2019), 'Academic on the blockchain', https://digitalcurrency.unic.ac.cy/free-introductory-mooc/self-verifiable-certificates-on-the-bitcoin-blockchain/academic-certificates-on-the-blockchain/

Wood, G. (2014), 'Ethereum: A secure decentralised generalised transaction ledger', Ethereum project yellow paper 151, 1 32.

Zheng, Z. et al. (2017), 'Blockchain challenges and opportunities: A survey (2017)', URL https://pdfs.semanticscholar.org/305e/dd92f237f8e0c583a809504dcec7e204d632.pdf pp. 6 7.

Manuscript submitted: 10.03.2019
Manuscript accepted: 04.06.2019

Address of authors:

Prof. Dr. Fabian Schär
Center for Innovative Finance
University of Basel
Faculty of Business and Economics
Peter Merian-Weg 6
CH-4002 Basel
Switzerland
E-Mail: f.schaer@unibas.ch

Fabian Mösli
BlockFactory Ltd
Eichstrasse 25
CH-8045 Zürich
Switzerland
Product Manager
E-Mail: fabian.moesli@blockfactory.com

Fabian Schär is Credit Suisse Asset Management (Switzerland) Professor for Distributed Ledger Technology and Fintech at the University of Basel, Switzerland and Managing Director of the University of Basel's Center for Innovative Finance.

BlockFactory's product manager Fabian Mösli developed the notarization platform certifaction.io with his team.