Towards Using Blockchain Technology to Prevent Diploma Fraud

Qiang Tang
Luxembourg Institute of Science and Technology
4362, Esch sur Alzette, Luxembourg
qiang.tang@list.lu
E-mail: qiang.tang@list.lu
Corresponding author: Qiang Tang (e-mail: qiang.tang@list.lu)

ABSTRACT After its debut with Bitcoin in 2009, Blockchain has attracted enormous attention and been used in many different applications as a trusted black box. Many applications focus on exploiting the Blockchain-native features (e.g. trust from consensus, and smart contracts) while paying less attention to the application-specific requirements. In this paper, we initiate a systematic study on the applications in the education and training sector, where Blockchain is leveraged to combat diploma fraud. We first present a general system structure for digitized diploma management systems and identify both functional and non-functional requirements. We then show that all existing Blockchain-based systems fall short in meeting these requirements. Inspired by the analysis, we propose a Blockchain-facilitated solution by leveraging some basic cryptographic primitives and data structures. Our analysis shows that the proposed solution respects all the identified requirements by design and can be further extended to enhance its security and privacy guarantees. Finally, we investigate the proposed solution’s computational complexity and demonstrate its practicality.

INDEX TERMS Diploma Fraud; Privacy; Security; Blockchain.

I. INTRODUCTION

In the education sector, diploma forgery is a long-standing problem [1]. With paper-based diplomas, forgery costs a little and it has been a prevalent business in many developing countries. Even in the developed countries where reputation is highly emphasized, diploma fraud is also not rare. It was reported in 2007 [2] that, Marilee Jones, the dean of admissions at the MIT, had fabricated her undergraduate degree and was forced to resign after working nearly three decades at the institute. The situation becomes better when Internet becomes ubiquitous and information sharing is made easy. However, until today, it remains difficult to validate diplomas. One of the major obstacles is the lacking of a universal platform between diploma issuers and diploma verifiers. In parallel to academic diplomas, fraud in industrial qualifications and working experiences is also a problem in the same vein.

To tackle the diploma forgery problem, many diploma verification solutions have been proposed and implemented. Most of them are institution-based, namely an institution offers an online system for diploma verifiers to validate users’ diplomas. Exceptionally, BADGR [3] and Open Badges [4] offer unified solutions for managing users’ entire educational history by collecting the digital certifications acquired by them at different academic institutes. Despite all the efforts, these solutions have not been widely used, and diploma fraud stays as a serious problem today.

A. BLOCKCHAIN IN A NUTSHELL

Following Nakamoto’s seminal work [5], the concept of Blockchain has become very popular in the society as it rapidly becomes the key enabling technology for the variety of cryptocurrency systems, including Bitcoin [5] and the altcoins. It is worth noting that Blockchain represents one special case of the distributed ledger technologies (DLTs), which are decentralised databases that rely on independent computers to record, share and synchronize digital
transactions. Regardless the different forms, a DLT promises similar properties to those from a Blockchain. For simplicity reasons, we choose to use the term Blockchain in this paper and our discussion generally applies to DLT.

Take the Bitcoin Blockchain as an example, repeatedly, a certain number of new data entries (e.g., transactions) will be packed into a new block by a miner and appended to the existing (longest) chain. The new block includes the hash value of the last block of the current chain, and is formed with some specific features, e.g., a proof of work (PoW) needs to be carried out so that the hash value of the new block contains some number of consecutive zeros. After its formation, the new block will be broadcast to the whole network, and it will be accepted by other network nodes after everything being validated. Blockchain systems act as the key foundation platform for smart contracts, which facilitate automated execution of software programs in a verifiable manner. One of the notable examples is Ethereum, which is the second largest cryptocurrency system after Bitcoin and gains the popularity because of its powerful smart contracts functionality.

Benefitting from its unique decentralised architecture, Blockchain possesses the following useful properties.

- **Decentralisation and Democracy.** Everyone can potentially act as a miner and has the same privilege to generate blocks and approve blocks to the Blockchain. This is generally true for systems employing the PoW as the consensus mechanism in the permissionless scenario, while it can be different in other cases. Regardless, Blockchain eliminates a single trusted party and anchors the trust base to multiple parties.

- **Integrity, Immutability and Consistency.** If an attacker or a group of colluded attackers does not dominate the consensus process, e.g., informally, in the case of Bitcoin Blockchain more than 51% of the computing power is at the hands of semi-honest miners, then it will not be able to modify the existing blocks that have been agreed on by the consensus. In other words, when being a majority, semi-honest miners can guarantee a consistent view for the Blockchain users and assure that no malicious attackers (including dishonest miners) can manipulate the blocks on the longest chain. Note that the trust assumption based on 51% semi-honest miners is only a theoretical bound, while it has been shown that 25% colluded miners can disrupt the operations of Bitcoin Blockchain [6].

- **Transparency, auditability and Disintermediation.** Comparing to existing information systems, Blockchain offers more transparency towards not only the data of blocks but also the origination of these blocks. In a permissionless Blockchain, everything is totally transparent to the world, while it is transparent to the authorized entities in other cases. Transparency naturally leads to auditability, and it also help eliminate many intermediaries in practice particularly when smart contracts feature is equipped.

Numerous Blockchain systems and applications have been proposed so far, and we refer the readers to the abundant surveys (e.g. [7]–[9]) and observatory reports for more information (e.g. those from the EU Blockchain Observatory & Forum).

### B. EMERGING BLOCKCHAIN-BASED SOLUTIONS

In the US, MIT is running a Digital Certificates project which uses Blockchain as a key infrastructure [10]. In EU, to support the digital single market, the Connecting Europe Facility (CEF) programme is funding a set of generic and reusable Digital Service Infrastructures (DSI). Among all, there exists a Blockchain DSI (the European Blockchain Services Infrastructure (EBSI)) which aims to accelerate the creation of cross-border services, where diploma validation is one of the selected use cases. Funded under EU’s Horizon 2020 research and innovation programme, the QualiChain project is dedicated to verifying educational credentials based on Blockchain [11] and the EDSSI project is a similar project [12]. In Cyprus, University of Nicosia [13] tried to digitize and decentralize their internal processes and have issued their first academic certificates as a proof of concept. In France, BCDiploma [14] shares the same goal towards a global certification network of higher academic institutions. In Switzerland, Gresch et al. [15] proposed a blockchain based system for managing diplomas called UZHBC (University of Zurich BlockChain), and Schär and Möсли [16] did a similar project through University of Basel’s Center for Innovative Finance and BlockFactory Ltd1.

Besides these initiatives, other researchers have also promoted Blockchain to solve the fraud issues in diplomas, e.g., Turkanović et al. [17], Serranito et al. [18], Tariq et al. [19], Brinkkemper [20]. Instead of focusing on technical solutions, Olivier et al. [21] investigated the business models for Blockchain-based diploma management solutions.

### C. CONTRIBUTION AND ORGANISATION

Most existing digitized diploma management systems have tried to directly transform the paper-based ancestors into digital systems. The aforementioned Blockchain-based systems move a step further to provide more guarantees on security and interoperability. However, a comprehensive modeling of digitized diploma management is still missing

1https://blockfactory.com/
today. In particular, security and privacy requirements have not been systematically studied.

In this paper, we close the gap by initiating a systematic study on digitized diploma management systems. Our main contribution is two-fold. Firstly, we present a general system structure and identify both the functional requirements (e.g. data included in a diploma and the time-stamping of diplomas) and the non-functional requirements from both the security and privacy perspectives. We also analyse the existing systems and show that they all fall short to meeting the identified requirements. Secondly, we follow the security/privacy-by-design principles to propose a Blockchain-facilitated diploma management solution. By relying on some basic cryptographic primitives (e.g. digital signature and hash function) and data structures (e.g. hash tree), we show that the proposed solution satisfies all the identified requirements and can be enhanced in several ways. The proposed solution is very efficient since there is no expensive data processing operations by the nature of diploma management.

The rest of this paper is organized as follows. In Section II, we describe the system architecture and identify the functional/security/privacy requirements. In Section III, we analyse existing (Blockchain-based) diploma management systems with regard to the identified requirements. In Section IV, we describe our new Blockchain-facilitated diploma management solution. In Section V, we present our analysis results of the proposed solution. In Section VI, we conclude the paper.

II. SYSTEM AND THREAT MODELING OF DIPLOMA MANAGEMENT

In a digitized diploma management system, the players fall into four categories including user (who is first a student and then becomes a graduate after being issued the diploma), diploma issuer, diploma verifier, and an intermediary platform. In our modeling, as shown in Figure 1, we assume there is one diploma issuer, which serves any number of users and diploma verifiers. The intermediary platform might be optional when the diploma issuer and the diploma verifiers directly interact with each other. Note however, in Section IV, we argue that we can solve the interoperability problem when diploma verifiers need to validate diplomas issued by different diploma issuers, by introducing Blockchain as the intermediary platform.

When setting up a digitized diploma management system, there should be an Initialisation Phase for all players to set up their parameters. The details of this procedure will depend on the specific solution, and we will not focus on it at this point. It is worth noting that this phase is not shown in Figure 1.

As indicated in the general system architecture in Figure 1, from the perspective of a user \( u \), the workflow consists of four phases. Note that the numbering (i.e. 1)-4) in the figure corresponds to the four phases below.

1) **Diploma generation.** User \( u \) registers and studies at an institute, which will act as the diploma issuer to issue a diploma to him once proper qualification is achieved. By default, we assume that user \( u \) gets a copy of all information related to himself, and stores it locally.

2) **Diploma outsourcing.** If an intermediary platform is employed, the diploma issuer stores some information about user \( u \)'s diploma on the platform, which will then take care of the following-up diploma verifications activities.

3) **Diploma usage.** User \( u \) presents some elements of his diploma to an organisation, who will then acts as the diploma verifier to validate these elements.

4) **Diploma verification.** The diploma verifier interacts with the intermediary platform to validate the elements received from user \( u \). If there is no platform employed, then the interaction is directly with the diploma issuer.

Clearly, the first two phases will happen only once (except that diploma may be updated), while the last two phases can happen as many times as user \( u \) wants. In the proposed architecture, we deliberately separate the diploma handling into two phases (i.e. Diploma usage and Diploma verification), to enforce our concept that no interaction between the user and the intermediary platform should be required for efficiency and facilitating other desired properties such as privacy.

Next, we elaborate on the functional and non-functional (i.e. security and privacy) requirements for a digitized diploma management system.

A. FUNCTIONAL REQUIREMENTS

Regarding the functionality, we would like to emphasize a major difference between paper-based diploma and the fully digitized one. In the
When designing a digitized diploma management system, we first need to consider the following threats which also exist its paper-based ancestor.

1) **Fake diploma issuer.** An attacker may pretend to be a legitimate diploma issuer and issues/sells diplomas for profits.

2) **Diploma forgery.** An attacker may try to impersonate the legitimate diploma issuer to forge diplomas on its behalf.

3) **Diploma issuer fraud.** A malicious diploma issuer can try to generate diplomas for users who are not officially registered. For example, the issuer can generate diplomas for non-existing users to exaggerate its performance, or can generate diplomas for celebrities to gain popularity.

Besides these, we need to tackle many new threats emerging in the digitized systems.

4) **Diploma issuer corruption.** When an attacker corrupts the diploma issuer and obtains its credentials, then this attacker may try to forge diplomas or do other harmful things (e.g. privacy leakages).

5) **User corruption.** When an attacker corrupts a user, as an immediate effect, it will result in privacy breaches. Besides, it may also cause security issues, e.g. blackmailing and denial of service.

6) **Intermediary platform corruption.** When an intermediary platform is employed, there is a risk that it will be corrupted. If this happens, the diploma verification service may be disrupted. It is worth noting that Blockchain possesses its own security vulnerabilities, therefore it is inappropriate to assume it to be corruption free.

7) **Diploma data confidentiality.** An attacker may try to obtain certain information in a user’s diploma while the user wants to keep such information confidential against the attacker. For our solution, the attacker refers to the diploma verifier who tries to learn information about the attributes that the user does not want to disclose.

8) **Diploma data integrity.** Once a diploma is issued, it should not be altered by any entity other than the diploma issuer. In case that any change needs to be made to the diploma, it should be done with the consents from the diploma issuer, the user and the time-stamping third party. For our solution, the attacker could be either the diploma issuer or the user. The time-stamping third party
is naturally assumed to be neutral and will not collude with any other party.

Note that there are other threats against digitized systems in general, e.g., denial of service attacks. We omit them in our discussion.

C. PRIVACY REQUIREMENTS

Privacy is not a serious concern in traditional paper-based systems due to the absence of effective information transmission media. However, things can change dramatically due to the introduction of intermediary platforms (e.g., Blockchain) and automated information processing and sharing capabilities (e.g., Blockchain’s smart contracts). It is worth emphasizing that automated information processing and sharing brings serious privacy concerns with respect to privacy regulations such as EU’s GDPR. Next, we categorize the privacy concerns as follows.

1) User privacy. Some user may want to hide the fact that his/her diploma has been verified. For example, if an employer notices that an employee’s diploma has been verified by its competitor, then it may infer that this employee has applied a job there. It is worth distinguishing this concern from the aforementioned diploma data confidentiality security requirement which enables a user to control the attributes in his/her diploma. For this privacy concern, the attacker refers to any entity other than the diploma verifier that the user is interacting with.

2) Diploma issuer privacy. In practice, a diploma issuer may choose not to reveal certain information about the issued diplomas, e.g., a small training center may not want to reveal the precise number of diplomas to the general public. As a remark, there might be some legal requirements on what the diploma issuer must reveal and what is optional. Addressing this privacy concern should align with the corresponding legal requirements. For this privacy concern, the attacker refers to any entity other than the users who naturally know the fact (e.g., the users who actually own these diplomas or the teachers).

3) Diploma verifier privacy. For a diploma verifier, its requests may reveal a lot of information, e.g., how many applicants it has, who are these applicants, when these applicants have submitted their applications, and so on. In practice, such information could reflect business secrets so that the diploma verifier may want to hide it as much as possible. For this privacy concern, the attacker refers to any other entity including the intermediary platform.

D. SUMMARY

In digitized diploma management systems, some traditional security concerns (e.g., fake diploma issuer, diploma forgery, and diploma issuer fraud) still exist, but it is much easier to address them with cryptographic tools such as digital signatures. On the other hand, many new security and privacy concerns emerge, partly due to the involvement of third-party intermediary platform. In particular, enriching diploma with more detailed data attributes amplifies the consequences of any security or privacy breach.

As mentioned above, combating fraud in paper-based systems is tedious and prohibitively costly, while it becomes much easier in digitized systems because this is the motivational design purpose. However, a digitized system introduces new cost, with respect to credential management, data storage and verification. Besides security and privacy, this aspect will affect the practicality of new digitized systems.

III. ANALYSIS OF EXISTING SYSTEMS

In this section, we analyse the existing digitized diploma management systems with respect to the identified requirements in Section II. Due to the fact that most of these solutions have been described at high level and focused on the workflow and smart contract design, our analysis stays at high level as well. Regarding the security and privacy requirements, no rigorous formulation and discussion have been provided for these solutions, so that our analysis results are only qualitative without in-depth cryptographic analysis.

A. MEETING THE FUNCTIONAL REQUIREMENTS

Table 1 indicates how well the existing solutions meet the data and time-stamping aspects of the functional requirements from Section II-A. For the former, most solutions only mention the term “diploma” (or, “certificate”) while providing no detail about it. In contrast, some solutions (e.g., [4], [10]) assume that any capability assertion can be contained in a diploma. For the latter, most solutions meet the requirement to some extent, benefiting from their adoption of Blockchain as the intermediary platform.

The efficiency and cost aspect is hard to analyse, and efficiency is often taken for granted as a result of employing smart contracts. From the perspective of business models, Olivier et al. [21] made some dedicated investigation from both the qualitative and quantitative aspects. They show that a sustainable business model relies on a lot of factors, such as market share, technology maturity, and acceptance of users and employers. Below, we emphasize three types of cost that have not been taken seriously in the existing systems.
C. MEETING THE PRIVACY REQUIREMENTS

Most existing solutions have security as their major objective, which is typically achieved by combining cryptographic primitives (e.g., digital signature) and Blockchain features. In contrast, privacy has been largely ignored and make these solutions vulnerable in reality. In Table 3, we indicate how the solutions meet the privacy requirements from Section II-C.

The solutions from [4], [15], [16] do not provide strong privacy guarantee because they employ permissionless Blockchain. It is worth noting that User privacy may not be guaranteed even though only the hash value of a diploma is stored on the Blockchain. The reason is simply because the hash value can uniquely identify the diploma and the user. In fact, EU’s GDPR has already pointed out this kind of vulnerability.

IV. NEW BLOCKCHAIN-FACILITATED SOLUTION

In this section, we first introduce the diploma format for our solution and then motivate the usage of Blockchain for diploma management. Finally, we describe our solution in detail.

A. DIPLOMA FORMAT

In order to support the selective disclosure of diploma attributes and avoid attribute-specific signatures, we propose to organise the attributes in a binary tree structure. For description simplicity, let’s assume that a diploma can include $N$ attributes where $N$ is an power of 2. We denote the attributes as $attr_1, attr_2, \cdots, attr_N$. For the sake of privacy protection as we will explain below, each attribute is accompanied with a salt value, denoted as $a-salt_i$ for $attr_i$. Let H be a cryptographic hash function, we can construct the attribute hash tree as follows.

1) Associate each leaf node with an attribute and a salt value, i.e. Node$_i$ is associated with $attr_i$ and $a-salt_i$. For each Node$_i$ (1 ≤ $i$ ≤ $N$), compute its value as $Hash[i] = H(attr_i || a-salt_i)$.

2) For internal node Node$[i|2]$ which has Node$_1$ and Node$_2$ as its children, compute its value as $Hash[1,2] = H(Hash[1] || Hash[2])$. Do the same for internal nodes Node$[3,4], Node[5,6], \cdots, Node[N-1,N]$.
For internal node Node_{[1,4]} which has Node_{[1,2]} and Node_{[3,4]} as its children, compute its value as \( H(Hash_{[1,2]}||Hash_{[3,4]}) \). Do the same for internal nodes Node_{[5,8]}, Node_{[9,12]}, \cdots, Node_{[N–3,N]}.

4) Continue as above until reaching the root node Node_{[1,N]} which has the value \( H(Hash_{[1,\frac{N}{2}]}||Hash_{[\frac{N}{2}+1,N]}) \).

For illustration purpose, this tree construction process is shown in a toy example with four attributes in Figure 2.

As most existing solutions do, it is natural to ask the diploma issuer to sign a user’s diploma. However, this is hardly sufficient to satisfy the desired security requirements in our threat model. For example, the diploma issuer can commit diploma issuer fraud without any barrier. Therefore, in our solution, we additionally require the user to sign the diploma as well. By doing so, the aforementioned diploma issuer’s fraudulent activities can be prevented. Suppose the user and the diploma issuer possess key pairs \( (PK_u, SK_u) \) and \( (PK_i, SK_i) \) respectively, for a secure digital signature scheme \( \text{Sign}(\text{Verify}) \). Then the user’s diploma will be in the following format.

\[
diploma_u = (\text{attr}_1, \text{salt}_1, \cdots, \text{attr}_n, \text{salt}_n, H(Hash_{[1,N]}),
\text{Sign}(H(Hash_{[1,N]}, SK_i)), \text{Sign}(H(Hash_{[1,N]}, SK_u)))
\]

For the sake of notation clarity, we use \( d_{\text{sign}} \) to denote a concatenation of the root value \( H(Hash_{[1,N]}) \) and the two signatures in the diploma \( \text{diploma}_u \). Namely,

\[
d_{\text{sign}} = H(Hash_{[1,N]}||\text{Sign}(H(Hash_{[1,N]}, SK_i))||\text{Sign}(H(Hash_{[1,N]}, SK_u)))
\]

It will be used to prepare information for Blockchain in Section IV-C.

### B. AN ATTEMPT WITHOUT INTERMEDIARY PLATFORM

Similar to paper-based diploma management solutions, a digitized solution can also be issuer-based or institution-based without any intermediary platform. In this case, the diploma issuer needs to deal with everything that is necessary for addressing every diploma verifier’s verification request. With this old-style design, we can easily observe the following challenges for the diploma issuer.

- The diploma issuer needs to be always online in order to deal with the potential request from any relevant diploma verifier. This stands for a high availability requirement, and furthermore implies a strong cyber-threat resilience requirement because the issuer is a lucrative cyber-attack target.
- The issuer needs to dedicate a lot of efforts to support diploma verification, e.g., maintaining a database for diplomas, deploying user interfaces for diploma verifiers, maintaining its signature key pairs and keeping a log of key update history.
- The issuer needs to leverage on some third party service to time-stamp diploma issuance. Note that the time-stamping operation is crucial to prevent fraudulent activities of the issuer itself.

Considering the diversity of all possible diploma issuers, it is to be expected that they will adopt various different systems of their own choices. These systems will have very different interfaces and workflows. It in turn poses a significant challenge to the diploma verifiers which are forced to use all these different systems.

### C. BLOCKCHAIN-FACILITATED SOLUTION

Based on our discussion in Section IV-B, it is clear that a (trusted) third party is required to provide the time-stamping service. Furthermore, it will be ideal if this third party can help address the challenges facing both the diploma issuer and diploma verifiers. In the following, we describe a solution by involving a Blockchain platform as the intermediary to provide the time-stamping service and facilitate other operations. An additional advantage of Blockchain is that it facilitate interoperability between different solutions which will be adopted by different diploma issuers.

---

**TABLE 2**

| Property                        | [16] | [19] | [17] | [18] | [15] | [4]  | [22] |
|---------------------------------|------|------|------|------|------|-----|------|
| Fake diploma issuer            | Yes  | Yes  | Yes  | Yes  | Yes  | ?   | Yes  |
| Diploma forgery                 | Yes  | Yes  | Yes  | Yes  | Yes  | ?   | Yes  |
| Diploma issuer fraud           | ?    | ?    | ?    | ?    | ?    | ?   | ?    |
| Diploma issuer corruption      | ?    | Yes’ | Yes’ | ?    | ?    | ?   | ?    |
| Intermediary platform corruption| ?    | Yes’ | Yes’ | ?    | ?    | ?   | ?    |
| Diploma data confidentiality   | Yes  | Yes  | Yes  | Yes  | Yes  | ?   | Yes  |
| Diploma data integrity         | Yes  | Yes  | Yes  | ?    | Yes  | Yes | Yes  |

**TABLE 3**

| Property                        | [16] | [19] | [17] | [18] | [15] | [4]  | [22] |
|---------------------------------|------|------|------|------|------|-----|------|
| User privacy                    | ?    | Yes  | Yes  | Yes  | ?    | ?   | Yes  |
| Diploma issuer privacy          | ?    | Yes’ | Yes’ | Yes’ | ?    | ?   | ?    |
| Verification requester privacy  | ?    | Yes’ | Yes’ | Yes’ | ?    | ?   | ?    |

---

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see [https://creativecommons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)
For instance, these solutions can use the same Blockchain platform (e.g. a consortium Blockchain operated by education agencies) and the same smart contracts. In this case, interaction with different solutions will be easy for the diploma verifiers.

We assume the Blockchain platform stays neutral from the diploma issuer and the users so that they will not be able to influence the standard operations of the platform. To this end, the diploma issuer or any user should not own any mining node of the Blockchain platform. For simplicity, we further assume that the Blockchain platform provides two smart contracts.

- **Contract\textsubscript{cert}** allows the diploma issuer and users to store, update and retrieve hash values of their public key certificates on the Blockchain platform.
- **Contract\textsubscript{diploma}** allows the diploma issuer to interact with the Blockchain platform, to store and update data on the platform. It also allows diploma verifiers to interact with the Blockchain platform to retrieve data.

The system structure of our Blockchain-based solution is shown in Figure 3.

**Initialisation Phase.** The diploma issuer and users should run the **Contract\textsubscript{cert}** smart contract to store hash values of their public key certificates on the platform, and also update these values when the signing keys have been revoked. As a remark, these entities should store their public key certificates locally. To interact with the Blockchain platform, every user should obtain a wallet which should allow him/her to securely store relevant data and execute smart contracts. Similarly, the diploma issuer should also obtain a wallet. At the same time, the issuer should maintain a local database to store data from all its users.

Referring to the general workflow from Section II, the detailed instantiation for our solution are as follows.

1) **Diploma generation.** After a user $u$ has passed the qualification tests, the diploma issuer issues diploma $\text{diploma}_u$ according to the format defined in (1). In addition, a salt value $d\text{-salt}_u$ is chosen to protect $\text{diploma}_u$. Use $u$ stores $\text{diploma}_u$ and $d\text{-salt}_u$ in his wallet, and the diploma issuer stores the same data in its local storage. In case the user’s wallet is lost, the diploma issuer can help him recover his data based on its local storage.

2) **Diploma outsourcing.** At a certain time, the diploma issuer batches the issued diplomas and store some information about them on the Blockchain as follows.

   a) The diploma issuer collects the diplomas for all relevant users. For simplicity, we assume there are $l$ diplomas and use the subscripts $1 \leq i \leq l$ to distinguish them. To hide the precise value of $l$, the diploma issuer chooses an integer $L \geq l$ and generates some fake diplomas $\text{diploma}_i$ for $l + 1 \leq i \leq L$.

   b) For each diploma $\text{diploma}_i$ ($1 \leq i \leq L$), the diploma issuer computes a hash value as
The functional requirements, from both the data aspect and the time-stamping aspect, have been addressed in Section IV-A.

c) Finally, the diploma issuer runs the Contract_diploma smart contract to store $\text{Hash}_i$ (1 ≤ i ≤ L) on the Blockchain. In addition, it also generates a signature for the concatenation of these hash values (i.e. $\text{Hash}_1||\cdots||\text{Hash}_L$) and stores it on the Blockchain.

3) Diploma usage. When the user u wants to present some attributes of his diploma to a diploma verifier, he should present the corresponding salt values for these attributes as well. In addition, he should also provide the necessary values of relevant leaf/internal nodes, so that the diploma verifier can compute the value of the root node of the attribute hash tree. Finally, he should provide the two signatures in his diploma. For instance, referring to the toy example Figure 2, if user u wants to present $\text{attr}_1$, then he should provide the following values:

$$(\text{attr}_1, a\text{-salt}_1, \text{Hash}[3,4], \text{Sign}(\text{Hash}[1,4], SK_I), \text{Sign}(\text{Hash}[1,4], SK_u))$$

4) Diploma verification. The diploma verifier runs the Contract_cert smart contract to retrieve the most up-to-date hash values of the public key certificates belonging to the diploma issuer and user u. Then, it performs some preliminary verification on the received data as follows.

a) It uses the received values from user u to construct the value of the root node for his diploma.

b) It then uses the retrieved public keys from the Blockchain to verify the received signatures from user u.

If the above verification passes, the diploma verifier runs Contract_diploma to retrieve the values stored by the the diploma issuer in step 2)c). Then, it continues the verification as follows.

c) It verifies the signature on the concatenation of hash values.

d) It finally checks that the diploma it receives from user u matches with one of these hash values.

The above two verification operations guarantee that the diploma information presented by user u is up-to-date (in case that the attributes in the diploma has been updated after its issuance) and also properly signed.

By design, the above solution has addressed the functional requirements, from both the data aspect and the time-stamping aspect. Moreover, we have minimized the amount of data on Blockchain and simplified the interaction between the diploma issuer/verifier and the Blockchain platform.

V. EVALUATION OF THE NEW SOLUTION

In this section, we first clarify the trust assumptions and some preliminary security notions for digital signature as well as hash function, and then perform our analysis.

A. ASSUMPTIONS AND PRELIMINARY BACKGROUND

For the security and privacy properties, we make the following standard trust assumptions, regarding the Blockchain, public key certification, and diploma issuer.

The Blockchain is neutral so that its operations will not be influenced by the diploma issuer and users. Some authority should regularly audit the Blockchain platform and its transactions resulted from the diploma management solution. Technically, auditing might be enforced using technologies such as those from [23]. It is worth noting that different types of Blockchain platform can be adopted for our solution.

- It is natural to assume that a group of higher education agencies can set up a consortium Blockchain and dedicate it to diploma management. The Blockchain can be further made permissioned so that access to the platform needs to be authorised.
- On the other hand, a permissionless Blockchain like Ethereum² can also work. From the legal perspective, we note that it might be difficult to regulate a permissionless Blockchain, particularly when it spans across different jurisdictions. Furthermore, the cost and efficiency aspects may be less predictable.

Different types of Blockchain have different security and privacy implications, particularly how much information will be shared with the general public. More details appear in the following-up analysis.

The public key certificates of the diploma issuer and users should be certified by respective authorities which act as the root of trust. How to determine these authorities is beyond the scope of this paper and will be decided when the proposed solution is actually deployed. We naturally assume that all entities have access to the faithful signature verification keys from the authorities.

The diploma issuer should faithfully carry out its duty. For example, it should properly manage its key pair: updating the key pair upon revocation and putting the hash value of the new public key certificate on the Blockchain immediately. It should stick to the workflow to generate diplomas and manage the...

²https://ethereum.org/en/
information on Blockchain. Except for these legitimate tasks, the diploma issuer may attempt to abuse the system for some additional benefit, e.g. committing a diploma issuer fraud.

For the signature scheme used in our solution, we require it to achieve the standard EUF-CMA (Existential Unforgeability under Chosen Message Attack) property. With this property, an attacker cannot forge a signature for a message \( m \) even if it can obtain signatures for any chosen messages different from \( m \). As to the hash function \( H \), we require it to be collision-resistant, which means that an attacker cannot find two different messages \( m_1, m_2 \) so that \( H(m_1) = H(m_2) \). Following the work of Bellare and Rogaway [24], sometimes we model the collision-resistant hash function as a random oracle in order to achieve rigorous proofs.

### B. SECURITY ANALYSIS

When designing our solution, we have adopted a security-by-design approach by taking into the security requirements from Section II-B. Next, we explain how these requirements have been addressed by our solution.

1) **Fake diploma issuer.** In our solution, the diploma issuer is required to possess a public key certificate certified by some authority and store its hash value on the Blockchain platform. Based on the EUF-CMA security property of the digital signature scheme, a fake diploma issuer will not be able to create a certificate on its own. Therefore, this threat is eliminated in our solution.

2) **Diploma forgery.** As long as the signature scheme is EUF-CMA secure, it will be computationally hard for an attacker to forge a diploma even if it can observe all existing diplomas.

3) **Diploma issuer fraud.** In our solution, a diploma needs to be signed by both the user and the diploma issuer, therefore it is computationally impossible for the diploma issuer to create a valid diploma for a user without his consent or signature. Therefore, the threat in this direction is eliminated. However, we note that if the diploma issuer colludes with users then it can issue seemingly legitimate diplomas. Management procedures like auditing can be employed to mitigate such fraud. We omit the details here.

4) **Diploma issuer corruption.** In our solution, when an attacker corrupts the diploma issuer and obtains its signing key, then this attacker can forge diplomas for users who collude with it. Since digest information of the diplomas is required to be stored on the Blockchain platform, then a forgery can be detected by the diploma issuer. Once detected, the forged diplomas can be made isolated by a combination of technical means, e.g. revoke the public key certificate so that the signature of the forged diploma becomes invalid and execute the Contract\( _\text{diploma} \) smart contract to overwrite the data from forged diplomas. Therefore, the threat of forging diplomas can be successfully eliminated in our solution. Besides forging diplomas, we do not see any other threats when the attacker obtains the signing key.

5) **User corruption.** In our solution, we have required every user to use a digital wallet for managing his diploma information. In reality, most of the wallets usually possess very strong security guarantees (e.g. designed based on trusted computing technologies) as they are often used to secure crypto-currencies. Therefore, this threat can be mitigated very well with existing products.

6) **Intermediary platform corruption.** We would like to point out that Blockchain platform is not a vulnerability-free tool that can resist all attacks. For a mission-critical application like diploma management, careful threat analysis should be carried out towards the chosen Blockchain platform. Next, we distinguish two cases.

- **In case of a consortium Blockchain platform,** the miners can be set to be the trusted entities in reality, e.g. the higher education agencies. Since the miners are well known, they can also be regularly audited to make sure that state-of-the-art cyber-security mechanisms are in place. By deploying a suitable consensus mechanism, the corruption risk can be made negligible given that the attacker cannot corrupt all the miners.

- **In case of a permissionless Blockchain,** the corruption risk becomes very tricky to analyse. On one hand, if the miner population is very big then it is practically difficult for the attacker to corrupt the majority of them. In addition, corruption may also be easy to detect. From this perspective, the corruption risk may be low. On the other hand, other related threats may arise. For example, the miners may unanimously boycott a certain diploma issuer and disrupt its business. Without a centralised governance mechanism, such threats are difficult to be addressed. In this case, the analysis should be done on a case by case basis, by going into the technical details of the target Blockchain platform.

Based on the above analysis, a consortium Blockchain platform is more preferable for our solution. Note that the Blockchain platform mainly offers the record keeping and time-stamping service in our solution. Based on this fact,
even if the platform is corrupted, the diploma management solution can recover easily if the diploma issuer can keep a local copy of the records on the Blockchain.

7) Diploma data confidentiality. In our solution, more precisely in the diploma data format definition from Section IV-A, salt values have been used for protecting attributes separately. In the Diploma usage phase, if a user does not reveal a-salt; then the diploma verifier cannot recover attr, even if it receives H(attr||a-salt). If we model H as a random oracle, then the verifier cannot distinguish attr, from any other value in the same domain.

8) Diploma data integrity. Due to the fact that the root value of the attribute hash tree is required to be signed by both the diploma issuer and the user, no modification can be made without the consents from both these entities based on the EUF-CMA security property of the underlying digital signature scheme.

C. PRIVACY ANALYSIS

In the process of designing our solution, we have tried to minimise the information disclosure to both the diploma verifier and the Blockchain platform, through organising the diploma attributes in a tree structure and integrating attribute-wise and diploma-wise salt values into the computation. Next, we show how the requirements from Section II-C have been addressed by our solution.

1) User privacy. For a user $u$, the only information stored on the Blockchain is $\text{Hash}_{u} = H(dig_{u}||d-salt_{u})$. The value of $d-salt_{u}$ is only known to user $u$, the diploma issuer and the diploma verifiers chosen by user $u$. Our solution offers two levels of privacy protection.

   a) If we model $H$ as a random oracle, then the attacker cannot distinguish $dig_{u}$ from any other value in the same domain, without the knowledge of $d-salt_{u}$. Therefore, it is computationally infeasible for the attacker to link $\text{Hash}_{u}$ to user $u$.

   b) Furthermore, in our solution, the diploma verifier will retrieve a batch of hash values to verify user $u$’s information, even if the attacker has learned the link between $dig_{u}$ and user $u$ (e.g. by colluding with a diploma verifier A who has received $d-salt_{u}$ from user $u$), then it is still impossible for it to tell which diploma another verifier is verifying.

2) Diploma issuer privacy. In our solution, by carefully choosing the value for $L$ in the Diploma outsourcing phase, the diploma issuer can hide the precise value of $L$. As we have mentioned before, the choice of value $L$ should align with the legal requirement (e.g. whether $L$ should be revealed to the public). In addition, there is a tradeoff between privacy protection and costs because increasing $L$ will increase the costs of data storage and smart contract execution.

3) Diploma verifier privacy. In our solution, the diploma issuer can enhance its privacy from several aspects. If the diploma verifier possesses multiple wallets, then it will be difficult for an attacker to link them. Moreover, the diploma verifier may also choose a trusted third party to interact with the Blockchain platform on its behalf.

As a remark, any information Revelation on permissionless Blockchain is visible to the general public, while consortium Blockchain is more privacy-friendly because the information Revelation is limited to the entities who can access the platform. It is fair to conclude that a consortium Blockchain is more preferable for our solution. Furthermore, the technologies from [25] might be integrated to improve the privacy guarantees in the implementation.

D. PERFORMANCE ANALYSIS

A full-fledged implementation of the proposed solution will take a long time, however this does not prevent us from evaluating the computational performance by implementing the cryptographic algorithms, namely hash function and digital signature schemes. Moreover, we also comment on the implementation of the smart contracts.

To benchmark the cryptographic primitives, we use a PC with an Intel® Core™ i7-4770 CPU @ 3.4 GHz processor with 16 GB RAM. For simplicity, we assume all entities has the same configuration for their PCs. We use SHA-256 to implement the function $H$ whose digest output has the length of 32 bytes, and summarize the performance results in Table 4.

| TABLE 4 HASHING PERFORMANCE |
|-----------------------------|
| Data item size (Kilobyte)   | 0.5 | 1  | 10 | 100 | 1000 |
| Hash one data item (ms)     | 0.004 | 0.007 | 0.068 | 0.689 | 6.698 |

For the digital signature scheme, according to NIST’s benchmarking [26], we choose Picnic [27] which can resist quantum attacks and is post-quantum secure. For the 128-bit security level, the signing takes about 10.34 ms and the verification takes about 2.49 ms. The signature has the length of 15 Kilobytes.

To estimate the computational costs for our solution, we assume there are 100 diplomas to be processed and each diploma has 16 attributes. We further assume each attribute can be expressed with 10 Kilobytes data and each salt value has the length of 32 bytes. We evaluate the cost for each phase and summarize the final results in Table 5.
1) Diploma generation. For each diploma, the diploma issuer needs 1.12 ms to prepare the diploma data and needs 10.34 ms to sign the diploma. For the user, it needs 10.34 ms to sign his diploma.

2) Diploma outsourcing. For simplicity, we assume the diploma issuer sets $L = \ell = 100$. In this case, the diploma issuer needs 68.9 ms to hash 100 diplomas, needs 0.068 ms to hash a concatenation of these hash values, and needs 10.34 ms to sign the data.

3) Diploma usage. We assume the communication between the user and diploma verifier will be performed off-chain, and the user’s computational cost is mainly from hashing the attributes that he does not want to disclose to the diploma verifier. This cost is upper-bounded by 1.12 ms.

4) Diploma verification. In this phase, the diploma verifier needs to compute the hash value for the root node of diploma attribute tree. This computational cost is upper bounded by 1.12 ms. Moreover, the diploma issuer needs to verify three signatures, by spending 7.47 ms.

From the figures in in Table 5, it is clear that our solution is very efficient and will be practical in reality.

For our solution, there are two smart contracts. Contract$_{cert}$ stores, updates or retrieves hash values of public key certificates on and from the Blockchain platform. Assuming the same setting as in the computational performance analysis, each hash value has the length of 32 bytes. Contract$_{diploma}$ stores, updates or retrieves hash values of diplomas on and from the Blockchain platform. In the same setting, the data has the length of 18 Kilobytes. Besides, there is no other business logic necessary in the smart contracts. Based on our security and privacy analysis from Section V-B and V-C, a consortium Blockchain will be a better choice in comparison to the permissionless ones like Ethereum. To this end, most consortium Blockchain platforms with smart contract features will be adequate for our purpose, e.g. Hyperledger fabric$^3$ or Corda$^4$ or Enterprise Ethereum$^5$. We leave it a future work to experiment with these platforms in the process of fully implementing our solution. In particular, we need to take into the specific business environment to evaluate the smart contract contract execution cost.

VI. CONCLUSION

In this paper, we have systematically studied digitized diploma management systems, from the system structure to requirements and to Blockchain-based solutions. The Blockchain-native record keeping and time-stamping features are crucial for our solution to prevent fraudulent activities from diploma issuers, while the integrity and immutability features make Blockchain an ideal platform to store diploma-related data and track key updating history. In comparison, the smart contract feature seems less important to us because we only need to store and retrieve data from the Blockchain. This leads to an observation that thinner variants of existing Blockchain platforms could suffice for our needs. It is an interesting future work to investigate this further and implement our solution for a detailed performance study.

REFERENCES

[1] H. Clifton, M. Chapman, and S. Cox, “Staggering trade in fake degrees revealed,” 2018. [Online]. Available: https://www.bbc.com/news/uk-42579634

[2] T. Levin, “Dean at M.I.T. Resigns, Ending a 28-Year Lie,” 2007. [Online]. Available: https://www.nytimes.com/2007/04/27/us/27mit.html

[3] I. Concentric Sky, “Badgr service,” 2021. [Online]. Available: https://info.badgr.com/

[4] I. G. L. C. Inc., “Open badges 2.x,” 2021. [Online]. Available: https://openbadges.org/

[5] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” https://bitcoin.org/bitcoin.pdf, 2008.

[6] I. Eyal and E. G. Siner, “Majority is not enough: Bitcoin mining is vulnerable,” in Financial Cryptography and Data Security - 18th International Conference, ser. LNCS, N. Christin and R. Safavi-Naini, Eds., vol. 8437. Springer, 2014, pp. 436–454.

[7] B. Butijn, D. A. Tambique, and W. Heuvel, “Blockchain: A systematic multivocal literature review,” ACM Comput. Surv., vol. 53, no. 3, 2020.

[8] M. Swan, Blockchain: Blueprint for a New Economy. O’Reilly, 2015.

[9] H. Treiblmaier and T. Clohessy, Blockchain and Distributed Ledger Technology Use Cases. Springer, Cham, 2020.

[10] D. C. Consortium, “Digital academic credentials,” 2021. [Online]. Available: https://certificates.media.mit.edu/

[11] Q. Consortium, “Decentralised qualifications’ verification and management for learner empowerment, education reengineering and public sector transformation,” 2021. [Online]. Available: https://qualichain-project.eu/

[12] E. Consortium, “EDSSI – European Digital Student Service Infrastructure,” 2021. [Online]. Available: https://edssi.eu/

[13] U. of Nicosia (UNIC), “Blockchain certificates (academic others),” 2021. [Online]. Available: https://www.unic.ac.cy/itl/blockchain-certificates/

[14] B. C. D. 2021, “Bcdiploma,” 2021. [Online]. Available: https://www.bcdiploma.com/en

[15] J. Gresch, B. Rodrigues, E. J. Scheid, S. S. Kanhere, and B. Stiller, “The proposal of a blockchain-based architecture for transparent certificate handling,” in Business Information Systems Workshops - BIS 2018 International Workshops, ser. Lecture Notes in Business Information Processing, W. Abramowicz and A. Paschke, Eds., vol. 339. Springer, 2018, pp. 185–196.

[16] F. M. F. Schar, “Blockchain diplomas: Using smart contracts to secure academic credentials,” Beiträge zur Hochschulforschung, vol. 41, no. 3, pp. 48–58, 2019.

3https://hyperledger-fabric.readthedocs.io/en/latest/smartcontract/smartcontract.html

4https://www.r3.com/corda-platform/

5https://entethalliance.github.io/client-spec/chainspec.html
[17] M. Turkanović, M. Holbí, K. Košič, M. Heričko, and A. Kamišalić, “Eductx: A blockchain-based higher education credit platform,” IEEE Access, vol. 6, pp. 5112–5127, 2018.

[18] D. Serranito, A. Vasconcelos, S. Guerreiro, and M. Correia, “Blockchain ecosystem for verifiable qualifications,” in 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services, BRAINS 2020. IEEE, 2020, pp. 192–199.

[19] A. Tariq, H. B. Haq, and S. T. Ali, “Cerberus: A blockchain-based accreditation and degree verification system,” 2019. [Online]. Available: http://arxiv.org/abs/1912.06812

[20] F. L. Brinkkemper, “Decentralized credential publication and verification: a method for issuing and verifying academic degrees with smart contracts,” June 2018. [Online]. Available: http://essay.utwente.nl/75199/

[21] M. Oliver, J. Moreno, G. Prieto, and D. Benítez, “Using blockchain as a tool for tracking and verification of official degrees: business model,” in 29th European Regional Conference of the International Telecommunications Society (ITS): “Towards a Digital Future: Turning Technology into Markets?”, 2018. [Online]. Available: http://hdl.handle.net/10419/184958

[22] D. C. Consortium, “Building the digital credential infrastructure for the future,” 2021. [Online]. Available: https://digitalcredentials.mit.edu/wp-content/uploads/2020/02/white-paper-building-digital-credential-infrastructure-future.pdf

[23] Y. Xu, J. Ren, Y. Zhang, C. Zhang, B. Shen, and Y. Zhang, “Blockchain empowered arbitrable data auditing scheme for network storage as a service,” IEEE Transactions on Services Computing, vol. 13, no. 2, pp. 289–300, 2020.

[24] M. Bellare and P. Rogaway, “Random oracles are practical: A paradigm for designing efficient protocols,” in Proceedings of the 1st ACM Conference on Computer and Communications Security. Association for Computing Machinery, 1993, p. 62–73.

[25] Y. Xu, C. Zhang, Q. Zeng, G. Wang, J. Ren, and Y. Zhang, “Blockchain-enabled accountability mechanism against information leakage in vertical industry services,” IEEE Transactions on Network Science and Engineering, vol. 8, no. 2, pp. 1202–1213, 2021.

[26] V. B. Dang, F. Farahmand, M. Andrzejczak, K. Mohajerani, D. T. Nguyen, and K. Gaj, “Implementation and benchmarking of round 2 candidates in the nist post-quantum cryptography standardization process using hardware and software/hardware co-design approaches,” Cryptology ePrint Archive: Report 2020/795, 2020.

[27] Microsoft, “The Picnic Signature Algorithm,” https://github.com/microsoft/Picnic.

QIANG TANG is currently a senior research scientist from Luxembourg Institute of Science and Technology (LIST). His research interests lie in applied cryptography, DLT/blockchain-enabled security design, and the privacy issues in machine learning. Dr. Tang received his Ph.D. degree from Royal Holloway, University of London, UK.