University of Zurich[UZH]

Department of Informatics

University of Zürich UZH
Binzmühlestrasse 14
CH-8050 Zürich
Switzerland

Prof. Dr. Burkhard Stiller
Christian Killer
Eder Scheid

Communication Systems Group
CSG@IFI
Head: Prof. Dr. Burkhard Stiller
Phone: +41 44 635 43 81(Supervisor)
Fax: +41 44 635 68 09
E-mail: stiller@ifi.uzh.ch
killer@ifi.uzh.ch
scheid@ifi.uzh.ch
URL:https://www.csg.uzh.ch/

Master Thesis (MA) for Simon Müller

Task Description:	Christian Killer, Eder Scheid, Prof. Dr. Burkhard Stiller
Title:	**Design and Implementation of a Data-Agnostic Structure for Blockchain Proof-of-Existence**
Start Date:	September 2, 2019
End Date:	March 2, 2020
Supervisor:	Christian Killer, Eder Scheid
Location:	Home, Communication Systems Group CSG
Main References:	[1] [2]

1. Introduction and Motivation

Academic certificates have major relevance in the labor market, signaling capability, and the level of education and skills of the recipient. Unfortunately, recent years have seen an increase in fraud, ranging from inflating academic grades to fake diplomas. Even several organizations focus on providing illegitimate academic degrees and diplomas (also called diploma mills). Estimating globally the number of individuals with fake diplomas is a hard task. In 2015, estimations indicated that about 41% of job applicants presented falsified information about their education in the US (United States) [13]. In 2017, it is estimated that about 500 fake doctoral diplomas are sold monthly in the US [14]. Thus, the release and verification of academic certificates is a known problem, tackled by academia [1], [2], [3], [5], and also private companies.

Public blockchains can be considered tamper-proof, transparent, without any centralized control, and they offer applications to a wide range of domains [2]. The main use-case applied to academic certificates is the Proof-of-Existence (PoE), *e.g.*, by first generating a unique cryptographic hash digest of a certificate and then publishing that hash to a public blockchain, effectively timestamping the certificate, thus, proving the existence of exactly this certificate, without leaking information about its content,typically the certificate's data. Recognizing the potential benefits of such a blockchain-based approach, prior work presented the necessary requirements for a solution at the University of Zurich (UZH) [2].

2. Related Work

Providing a trustworthy, decentralized, and publicly available data storage, public blockchains have become a disruptive technology that has seen interest across academia and industries alike. Many interesting projects (blockchain-based or not) have explored the possibility to digitally verify diplomas to counteract the trend of fake degrees.

Blockcerts [3],[4] is an initiative by the MIT (Massachusetts Institute of Technology) to create an open standard for issuing and verifying credentials on the Bitcoin blockchain. The system is now in use at MIT [4] and empowers graduates to use the service through a mobile app [15]. Similar to that approach, the National Research and Education Network of Greece (GRNET) [16] also persist diploma hashes to a public blockchain. However, the GRNET project [16] differs from Blockcerts because not only hashes of diplomas can be stored, but also the entire verification process. Therefore, verification requests, successful or unsuccessful proof and the forwarding of the result to its requester are steps that will be stored. Another mentionable initiative is led by the Trust::Data Consortium [17] from MIT, aiming to provide safe distributed computation, enabling privacy-preserving data sharing [17]. Further, UNIC (University of Nicosia) [5] initiated a blockchain-based project to issue and verify academic certificates. UNIC aims to digitize and decentralize their internal processes issuing their first academic certificates as a Proof-of-Concept (PoC).

Generally, the same approach can be found in almost all related and blockchain-based work of academic certification. Most projects only persist the hash of the certificate into the public blockchain, while the certificate data is then sent to the recipient, who can share it with others, such as an employer. These credentials can be used to create the same fingerprint that can be found in the blockchain and, thus, verify its veracity. The amount of related work tackling the problem of academic certification highlights its necessity.

3. Description of Work

The focus of this Master Thesis is the design and implementation of a data structure that is extensible and usable in blockchain-based certificate issuance process [1]. While the initial use-case scope entails digital diplomas (certificates), the data structure cannot be restricted to academic certificates, but need to be extensible to other structured data, especially a Curriculum Vitae (CV) or recommendation letters, scientific publications, attributions of attendance, including all forms of publicly verifiable data.

In the first stage, the student is required to research relevant related work for suitable data structures and research work within academia and relevant certificate standards. Also, the first stage includes the evaluation and possibilities to integrate existing legacy systems and data structures with the new SwissEduchain [1].

The second stage is concerned with the design and implementation of an application that can be used by the Recipient (*e.g.*, a student) and the Verifier (*e.g.*, an employer). The architecture shall be discussed during periodical meetings with the advisor to examine the feasibility of the proposal; additionally a close contact shall be maintaned continuously with the related Master Thesis on “Identity Management for a Blockchain-based Certificate Issuance”. The expected outcome is a working PoC(Proof-of-Concept) that adheres to the designed solution with a detailed description and reasoning of implementation decisions taken.

The final stage of this Master Thesis covers an evaluation with respect to its achieved properties and a discussion of the implemented PoC. These results need to be contrasted to the thesis goals. Also, decisions on how to best evaluate the PoC must be made in close synchronization meetings with the supervisors. A final report must include a motivation and problem description, background information, related work, design decisions, implementation details, evaluation, and conclusions (see below for the formal requirements). Note that an initial version of the report is required at the time of the midterm presentation describing the current status of the thesis.

It is important to note that this work presents a strong exploratory aspect, since the detailed implementation challenges and practical obstacles of designing and implementing a blockchain-based SwissEduchain system are still unknown. Thus, adaptations to the project goals may occur during the development of this Master Thesis. The main success indicator of the thesis are the successful deployment and demo of an end-to-end issuance, receival, and verification process.

4. Master Project Goals

The main goal of this Master's Thesis is to design and implement an extensible data structure that can be used in the SwissEduchain project [1] with an application used by the Recipient and the Verifier. Driven by the description of work outlined, the following key goals for this thesis are determined:

- **Investigate the suitability of multiple Blockchain platforms for blockchain-based certificate verification process:** The comparison and evaluation of different public [6],[7] and private [8],[10] blockchain or DLT (Distributed Ledger Technology) platforms should be documented and support the decision for a specific platform and architecture.
- **Requirements Engineering:** Elicit requirements by evaluating the current process of certificate verification process (How and by whom are they exactly produced and stored? [1],[2]). Document the current process and propose improvements with the new Swiss Educhain design.
- **Research the individual requirements with regards to Privacy, Security and Verifiability:** With regard to the publication of hashes, evaluate from a privacy, security and verifiability perspective, evaluate potential risks and problems, but also advantages.
- **Design and Architecture:** Design a flexible data structure and applications with a good user experience (UX) in mind. Create an architecture fulfilling prior defined properties (*e.g.*, data structure should be extensible, modular).
- **PoC Evaluation:** Evaluate the implemented approach considering its prior defined properties.
- **Code Delivery and Testing:** Source code needs to be well-documented, open-source and readable. The PoC is be tested with appropriate methods.
- **Documentation and Report:** The steps of the initial analysis, its results, design decisions, prototyping, and the evaluation approach and its results are to be documented in the final thesis report.

Furthermore, these goals may result in slightly different requirements, which may influence the activities scheduled. In this sense, premises could be considered in agreement with the supervisor in order to restrict, change, or adapt activities to the scope of the project, and consequently, its objectives. Moreover, a scientific publication containing the results of the thesis might be produced. Thus, the development of the publication must be aligned with the project development.

5. Activities and Milestones

Based on the thesis description, the following activities need to be accomplished during the independent study.

- Planning: discuss initial planning with the supervisors, defining the thesis scope and requirements. Thus, it involves the specification of tasks and responsibilities, as well as an initial time estimative to conclude each planned task of the thesis.
- Theoretical Basis: include a detailed overview of techniques and solutions towards the proposal. This milestone must reveal the complexity involved in the design, development, and integration of each task into the proposed system, serving as the basis to the design and prototype development milestones.
- Design: based on the elements of the theoretical basis involved in the proposal, the design may be readjusted to match the code complexity and project timing. This milestone must reveal the architectural elements and its requirements towards the thesis goals.

- Prototype Development: comprises the implementation and integration of architectural elements described in the design milestone. Also, this stage must output results and material to be contrasted with the prototype requirements and thesis goals.
- Evaluation: run prototypical experiments to produce data to be contrasted with thesis goals and requirements. Also, this milestone involves the specification of evaluation scenarios and use cases, establishing parameters for the collection of data and further comparative analysis based on produced results.
- Documentation and Report: the report must describe all the activities and milestones being, ideally, produced in parallel with the prototype development. Thus, providing enough time so as the supervisors could perform partial reviews.

6. General Notes

- **Schedule/Time Planning**: the student has to provide a written schedule for his/her full study steps within the first two weeks of his/her work. Clarify details with your supervisor and finalize the schedule of tasks, basically in a weekly fashion. Include the intermediate/public presentation(s), too.
- **Intermediate Presentation**: prepare an intermediate and internal presentation (20 min max plus Q&A) after half time of your study (date to be set) and discuss this with your supervisor. At the end (date to be set) a final public and self-containing presentation (20 min max plus Q&A) has to be given.
- **Report**: final written report will document all work undertaken and remember that this report must be self-contained. Major technical basics are to be included and detailed knowledge obtained during the work must be documented. Assumptions, design decisions, configuration choices, and results are part of the report as well as design details and usage information. Correct bibliographic references and a list of papers, recommendations, and descriptions used must be added, including those ones given below.
- **Period Meetings with Supervisors**: establish periodic meetings with the supervisor to report on progress, discuss current problems and request assistance when required.
- **Interaction**: students involved in this proposal must promptly read and answer emails related to this project. If needed, a more frequent interaction will determine a better basis for supervision and progress.

7. Formal Results

Besides the intermediate and final oral presentations, the following documents need to be handed-in to the supervisor before the final deadline:

- **Report Printed Copies**: double-sided report in a soft cover binding and in 3 copies (in English or German): the report must cover the milestones, a table of content and figures (including tables), a valid list of bibliographic references, and optional appendices as required is part of the report, too. The official acknowledgement section is mandatory, a personal one optional, however recommended, as usually a number of people took part in the process of finalizing the study. The text processing shall be done in LaTeX (preferred) or FrameMaker and in rare cases in Word.
- **Source-Code and Documentation**: A documentation of the design/configured system, covering the system's view point, a description, a use and installation manual, a documentation of all program and data structures developed or utilized is essential. All of this may be part of the report above, however, in case it will not be included, it must exist in a separate document.
- **Digital Copy**: a dedicated CD has to be produced containing (1) a collection of all program sources, software components, protocol design utilized, and documentation in PDF; (2) the written report in PDF or full HTML pages, a directory description if needed, figures in source file and JPG or EPS and a full printable PS as well as PDF file; (3) the set of slides for the final presentation in

PPT (PowerPoint); and (4) all further material used, if available in electronic form, such as all existing and documented tested scenarios, plans, and results.

- **German Summary**: maximum of 2 pages (in case the report is written in German, the summary needs to be in English), which will enable a quick and clear overview of the work, tasks and results. The summary will be part of the bound report, and is included after the front page and before any other following text. It includes four short sections: 1. Einleitung/Introduction, 2.Ziele/ Aims and Goals, 3. Resultate/Results, and 4. Weitere Arbeiten/Further Work.
- **Hand-in**: The complete set of copies of the report, the CD, and the talk must be completed and handed in to the supervisor in time before the study will change into the "submitted" status. Note that failures of delivering those information and data after the formal hand-in time may result in a “non-passing” state of this work.

8. References

[1] Christian Killer, Eder Scheid, Bruno Rodrigues, Geetha Parangi, Burkhard Stiller: Swiss EDUCHAIN Project Proposal. Project Proposal. Universität Zürich, Communication Systems Group, Department of Informatics, Zürich, Switzerland, Juli 2019.

[2] Jerinas Gresch, Bruno Rodrigues, Eder Scheid, Sali Kanhere, Burkhard Stiller: The Proposal of a Blockchain-based Architecture for Transparent Certificate Handling. Business Information Systems (BIS) 2018 International Workshops, Berlin, Germany, July 18, 2018, Revised Papers, pp.185-196.

[3] MIT Registrar’s Office. Digital Diploma. Available [Online] http://bit.ly/mit-registrar, Accessed August 12, 2019

[4] MIT Registrar’s Office. Digital Diploma Program FAQs. Available [Online] https://bit.ly/2JYw4zT, Accessed August 12, 2019

[5] University of Nicosia. Academic certificates on the Blockchain. Available [Online]: https://bit.ly/2I5G3mj, Accessed August 12, 2019

[6] Gavin Wood. Polkdaot: Vision for a Heterogeneous Multi-Chain Framework. November 2016, Available [Online]: https://polkadot.network/PolkaDotPaper.pdf, Accessed August 12, 2019

[7] Ethereum, Available at https://ethereum.org, Accessed August 12, 2019

[8] Corda, Available at https://corda.net, Accessed August 12, 2019

[9] Hyperledger Wiki. Accessed Aug 15, 2019. Available [Online]: https://wiki.HL.org/

[10] Hyperledger Fabric Wiki. Accessed Aug 15, 2019. Available [Online]: https://wiki.hyperledger.org/display/fabric/

[11] Hyperledger Indy Wiki. Accessed Aug 15, 2019. Available [Online]: https://wiki.hyperledger.org/display/indy/

[12] Sovrin Foundation. Accessed Aug 15, 2019. Available [Online]: https://sovrin.org/

[13] N. M. Musee, An Academic Certification Verification System Based on Cloud Computing Environment. PhD dissertation, University of Nairobi, 2015

[14] H. Park and A. Craddock. Diploma Mills: 9 Strategies for Tackling One of Higher Education's MostWicked Problems. Accessed Aug 15, 2019. Available [Online]: https://bit.ly/2DoEeyu

[15] Google Play Store. Blockcerts Wallet. Accessed Aug 15, 2019. Available [Online]: http://bit.ly/blockcerts-wallet

[16] A. Castor. Cardano Blockchain's First Use Case: Proof of University Diplomas in Greece. Accessed Aug 15, 2019. Available [Online]: https://bit.ly/2DVsrYt

[17] Trust::Data Consortium - An initiative of MIT Connection Science. Accessed Aug 15, 2019. Available [Online]: https://www.trust.mit.edu/about

Zürich, ~~July 1~~August 28, 2019

Prof. Dr. Burkhard Stiller