frontiers

Integrating bloxberg's Proof of Existence service with MATLAB

Kevin Wittek [1,*], Dominik Krakau [1], Neslihan Wittek [2], James Lawton [3] and Norbert Pohlmann [1]

[1] *Blockchain Research Lab, Institute for Internet Security, Westphalian University of Applied Sciences, Gelsenkirchen, Germany*
[2] *Department of Biopsychology, Faculty of Psychology, Ruhr University Bochum, Bochum , Germany*
[3] *Digital Labs, Max Planck Digital Library, Munich, Germany*

Correspondence*:
Institute for Internet Security, Westphalian University of Applied Sciences,
Neidenburger Str. 43, 45897 Gelsenkirchen, Germany
wittek@internet-sicherheit.de

ABSTRACT

Proof of Existence as a blockchain service has first been published in 2013 as a public notary service on the Bitcoin network and can be used to verify the existence of a particular file in a specific point of time without sharing the file or its content itself. This service is also available on the Ethereum based bloxberg network, a decentralized research infrastructure that is governed, operated and developed by an international consortium of research facilities. Since it is desirable to integrate the creation of this proof tightly into the research workflow, namely the acquisition and processing of research data, we show a simple to integrate MATLAB extension based solution with the concept being applicable to other programming languages and environments as well.

Keywords: blockchain, ethereum, poe, poa, bloxberg, dlt, open science

1 INTRODUCTION

Researchers predicate their work on the earlier findings of their topics and investigate the remained questions to go further. However, for progress, they have to cope not only with their experimental design and methodology but also the replicability of earlier findings. The replicability crisis has been highlighted in the last decade by affecting the vital research areas like social sciences, natural sciences and medicine (Pashler and Wagenmakers (2012)). The survey that was conducted with over 1500 scientists revealed that almost three-quarters of them failed to reproduce another scientist's experiment and curiously enough, half of them even collapsed to reproduce their own experiments (Baker (2016)).

There are several reasons behind this crisis, but the first and foremost is the pressure on the researchers' shoulders to publish in a credible journal. This pressure causes p-value hacking by applying many statistical tests on the data and reporting only the significant and mostly positive results which do not represent the real findings. The logical way to overcome these obstacles is to be more open about the published research by sharing the detailed explanation of the experimental design, methodology and analyzed data, as well as the unretouched raw data.

Miyakawa has reported that after asking several authors to provide raw data as an editor in chief of a journal, most of the manuscripts were withdrawn or the raw data was not sufficient enough to be published as of its written (Miyakawa (2020)). All these deceptions induce a waste of time and resources for the scientists who are willing to investigate more about their research interest and demonstrate both positive and negative results. Now we are at a point in time in which people are aware of the issues mentioned above and instead of only discussing this on media, researchers should shoulder responsibility and take a step to encounter the replicability crisis by being transparent on raw data sharing.

In this context, the blockchain and distributed ledger technology (DLT) might be an enabling factor to allow for a better digitalization and automation, leading to an improvement in integrity and transparency of research data and the research process as a whole. A Bitcoin-based implementation of a Proof-of-Existence (PoE) service for generic digital documents has already been released as early as 2013 and uses the approach of storing a cryptographic document hash on the public ledger. It, therefore, acts as a public notary service for proving the existence of a document at a certain point in time without disclosing the content of the document itself (Swan (2015); Kirk (2013)). A similar approach has been proposed for a secure and tamper-proof storage of clinical trial data, hinting at the potential for improving the general quality of clinical research with regards to traceability, prevention of a posteriori reconstruction of data and secure automation (Benchoufi and Ravaud (2017)). In addition, blockchain and DLT based solutions have been suggested for solving problems in the intellectual property and copyright domain, for example, for a secure timestamped manuscript submission and peer review system (Gipp et al. (2017)) and a traceable collaborative design thinking and open innovation platform (Schönhals et al. (2018)).

Based on these findings and the demands of the scientific community, we have developed a software library for the integration of the industry-standard MATLAB computing environment with the scientific blockchain infrastructure bloxberg, which allows for seamless inclusion of raw research data existence certification into existing scientific processes.

2 BLOXBERG

The bloxberg infrastructure is a secure global blockchain governed and secured by an international consortium of scientific organizations. The infrastructure's goal is to provide scientists with services based on blockchain as well as fostering collaboration between the scientific community (Kleinfercher et al. (2020)).

In comparison to other prominent blockchain networks such as Bitcoin and the public Ethereum network (Mainnet) which utilize Proof-of-Work (PoW) (Nakamoto (2008); Wood (2019)), the bloxberg blockchain uses a Proof-of-Authority (PoA) consensus engine with Authority Round (Aura) as the used consensus algorithm (Kleinfercher et al. (2020)). This algorithm minimizes the energy consumption of securing a blockchain and increases the potential throughput while maintaining decentralization by distributing block confirmations between the participating scientific organizations (Parity (2020)). However, since Aura relies on UNIX time synchronization of authority nodes, situations might occur, where different sets of authorities have a different current leader, leading to concurrent forks of the chain, which are resolved eventually over time (Angelis et al. (2018)). As a consequence, this means bloxberg has no-consistency or eventual-consistency guarantees, making it an AP (availability, partition tolerance) system in the context of the CAP-theorem.

Furthermore, since the validating nodes in the network are known entities, specific computational and network requirements can be met by participating nodes. This property ensures a higher degree of scalability and efficiency compared to PoW-based blockchains, while at the same time implementing the concept of a based distributed trust architecture in the form of a consortium of international research organizations.

These properties make the bloxberg network an ideal infrastructure to build scientifically-focused blockchain applications on. One of those already existing applications is the Certify DApp.

3 CERTIFY DAPP

The Certify DApp is a production-ready decentralized application deployed on the bloxberg network. It can be used to verify the existence of an arbitrary file (i.e., generic research data) at a certain point in time without disclosing the content of the file itself.

This use case is implemented by recording the SHA-256 hash, which is considered a strong cryptographic hash function by the German Federal Office for Information Security (BSI (2019)), together with additional metadata as a transaction in the bloxberg network (see figure 1). The transaction creation timestamp acts as a public record, proving the existence of the certified data at this point in time. It is, therefore, possible to later verify the prior certification of a file by looking up the timestamp of the first transaction containing its SHA-256 hash in the bloxberg network.

In addition to being usable as a real DApp in conjunction with a wallet software (e.g., MetaMask), it is also possible to interact with the Certify DApp as with a regular web application. This access is implemented utilizing a web application, including a REST-API, which interacts as a proxy, or intermediary agent, towards the bloxberg network. Additionally, further accessibility and user experience improving clients and integrations are possible, such as Max Planck Digital Library's single-button integration into their existing internal cloud storage solution KEEPER (MPG (2019)).

4 SYSTEM DESIGN

Our system design tries to strike a compromise between accessibility (with scientists in general as the intended user group) and leveraging the capabilities of a distributed trust architecture, while at the same time being easily integratable into existing scientific processes. We, therefore, opted for an integration via a single MATLAB file, which can be added to any existing MATLAB project without requiring the installation or configuration of any external components.

It is implemented in an object-oriented design as a MATLAB class. Since MATLAB brings interoperability features with Java out of the box, the implementation occurs as Java code interwoven with the MATLAB class structure (see supplementary material). Since the implementation is using the existing Certify DApp web service, no key management and wallet software is necessary and the communication between MATLAB and the web service occurs over HTTP, with the Certify DApp web service acting as a proxy to the bloxberg network (see Figure 2). This of course means, that this design accepts the Certify DApp web service as a trusted component for this process. While this increases the accessibility and lowers the entry bar, it, at the same time, introduces a single point of failure into the system and works against the data autonomy. Future implementations might expand on this problem.

The provided API allows the certification of any file from within the program flow (see listing 1). It is, therefore, possible to certify final as well as intermediary results alike. This API might be extended in the future to also allow the certification of generic MATLAB data structures. However, note that the hash-based approach of certification lends itself better to certifying persistent artifacts in order to allow for the creation of the relation between the certified hash and the actual research data.

5 DISCUSSION AND FUTURE WORK

The current implementation does not make use of any decentralized public-key cryptography infrastructure and therefore does not provide strong guarantees with regards to the origin of the certified research data. A substantial improvement would be the integration of key management in the form of wallet software. This component would allow cutting out the Certify DApp web service as a middle-man and would, therefore, lead to a real DApp implementation.

Also, the current approach allows solely for the certification of single pieces of research data as an atomic unit. A much more significant potential lies in the possibility of certifying the scientific process as a whole over its complete lifetime. This concept might include entities such as experimental designs and methodologies, experimental setups, used hardware (ideally in the form of cyber-physical systems), source code and used software, experiment subjects (e.g., digital identities of humans and animals) and experiment conductors in addition to the intermediary and final research data and results.

Current efforts of the bloxberg community in the form of bloxberg Improvement Proposals (BLIPs) already try to tackle this challenge of certifying a multi-dimensional scientific process (bloxberg (2020)). The particular work item for this is BLIP-0001, Research Object Certification. This BLIP process is modeled after established community based software standartization efforts such as Ethereum Improvement Proposals (EIP) (Ethereum (2020)) and JDK Enhancement Proposals (JEP) (Reinhold (2020)).

All corresponding source code is published under the MIT open source license on GitHub[1].

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

NW contributed the introduction and proposed a general scientific outlook on the topic. JL contributed substantial parts about bloxberg and the Certify DApp, as well as figure 1. Software implementation was contributed by DK. NP provided overall supervision. KW contributed the conception, design and software architecture as well as the overall writing of this document. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was partially supported by the Ministry of Economic Affairs, Innovation, Digitalisation and Energy of the State of North Rhine-Westphalia as part of the connect.emscherlippe project at the Westphalian University of Applied Sciences in Gelsenkirchen.

ACKNOWLEDGMENTS

We would like to thank all members and institutions of the bloxberg community for their ongoing support and dedication.

[1] https://github.com/internet-sicherheit/bloxberg-matlab

REFERENCES

Angelis, S. D., Aniello, L., Baldoni, R., Lombardi, F., Margheri, A., and Sassone, V. (2018). Pbft vs proof-of-authority: applying the cap theorem to permissioned blockchain. In *Italian Conference on Cyber Security (06/02/18)*

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature News* 533, 452. doi:10.1038/533452a

Benchoufi, M. and Ravaud, P. (2017). Blockchain technology for improving clinical research quality. *Trials* 18, 335. doi:10.1186/s13063-017-2035-z

bloxberg (2020). bloxberg-org/blips

BSI (2019). BSI – Technische Richtlinie - Kryptographische Verfahren: Empfehlungen und Schlüssellängen

Ethereum (2020). Ethereum Improvement Proposals. *Ethereum Improvement Proposals*

Gipp, B., Breitinger, C., Meuschke, N., and Beel, J. (2017). CryptSubmit: Introducing Securely Timestamped Manuscript Submission and Peer Review Feedback Using the Blockchain. In *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. 1–4. doi:10.1109/JCDL.2017.7991588

Kirk, J. (2013). Could the Bitcoin network be used as an ultrasecure notary service? *Computerworld*

Kleinfercher, F., Vengadasalam, S., and Lawton, J. (2020). bloxberg - The Trusted Research Infrastrucutre - Whitepaper 1.1

Miyakawa, T. (2020). No raw data, no science: another possible source of the reproducibility crisis. *Molecular Brain* 13. doi:10.1186/s13041-020-0552-2

MPG (2019). First international blockchain for science: bloxberg

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System

Parity (2020). Parity Documentation - Aura - Authority Round

Pashler, H. and Wagenmakers, E.-J. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 7, 528–530. doi:10.1177/1745691612465253

Reinhold, M. (2020). JEP 0: JEP Index

Schönhals, A., Hepp, T., and Gipp, B. (2018). Design Thinking using the Blockchain: Enable Traceability of Intellectual Property in Problem-Solving Processes for Open Innovation. In *Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems - CryBlock'18* (Munich, Germany: ACM Press), 105–110. doi:10.1145/3211933.3211952

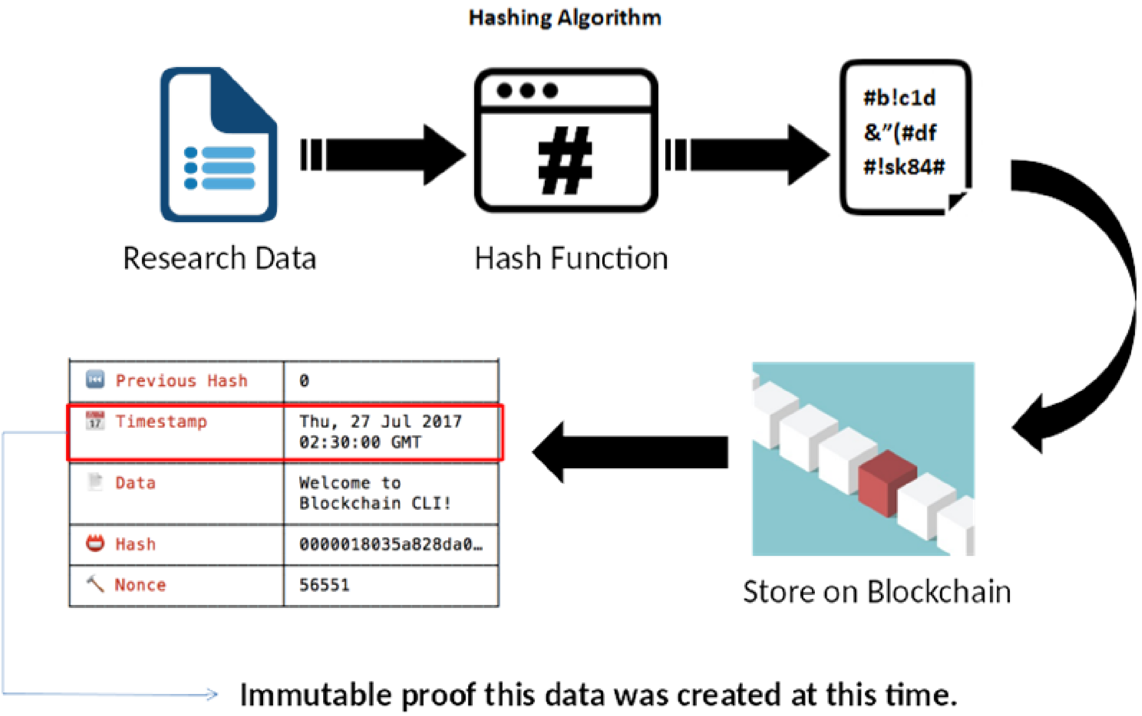
Swan, M. (2015). *Blockchain: blueprint for a new economy* (Beijing : Sebastopol, CA: O'Reilly), first edition edn. OCLC: ocn898924255

Wood, G. (2019). Ethereum: A secure decentralised generalised transaction ledger byzantium version 7e819ec

FIGURE CAPTIONS

```
MBB = MatlabBloxbergAPI('John Doe', 51200,
↪  'https://certify.bloxberg.org/certifyData',
↪  'https://certify.bloxberg.org/generateCertificate');
MBB = certifyData(MBB, 'researchdata.mat');
generateCertificate(MBB, 'C:\Users\John\Desktop',
↪  'mycertificate.pdf');
```

Listing 1 Example API usage



Previous Hash	0
Timestamp	Thu, 27 Jul 2017 02:30:00 GMT
Data	Welcome to Blockchain CLI!
Hash	0000018035a828da0…
Nonce	56551

Figure 1. Research data certification process in bloxberg Certify DApp

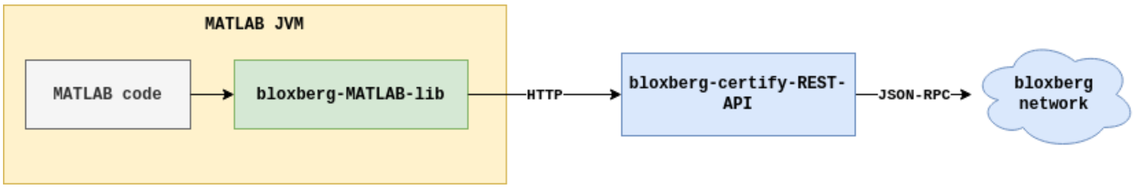


Figure 2. Components, relations and protocols