4H1-OS-9a-02

The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Connecting Learning Histories of Students Across Different Institutions Using Blockchain

Hiroaki Ogata[*1] Patrick Ocheja[*2] Brendan Flanagan[*1]

[*1] Academic Center for Computing and Media Studies, Kyoto University. [*2] Graduate School of Informatics, Kyoto University

Abstract

Learning records represent the activities of learners on a learning platform. Because learning takes place at different institutions, organizations and/or platforms, it is important to connect learning records belonging to the same learner on these various platforms for a wider spectrum of analytics. With decentralization at the heart of the blockchain technology, we show the implementation of a blockchain based learning analytics platform. By using smart contracts, we enforce restricted access to learner's data and empower learners with more control over their learning records. To ensure that learning records are immutable, we use a hashing strategy to detect changes between earliest version of a learning record and subsequently retrieved versions.

1. Introduction

Learning histories refer to the learning activities performed by students on learning platforms. While this information may vary across platforms and institutions, learning histories often give more information about a student's learning behaviors and outcome. Such information may include time spent on particular lesson modules, memos written, questions asked, results of quizzes and assignments, grades obtained and more. Unfortunately, upon graduating from a particular institution and moving to another (e.g. graduation from undergraduate program to a graduate program in a different university), students only take with them a certificate of completion and transcript of scores. While such certificates provide a summary of students' accomplishments, they do not provide a robust foundation for learning personalization on learning platforms at their new university of study. It is on this premise that we propose that learning histories of students should be transferable from one institution to another without hitches.

Although there are many institutions and many learning platforms built with different technologies, standardizations such as Tin Can Experience API [1] and IMS Caliper Discovery API [2] provide a uniform way of defining learning histories or records independent of the institution or platform. Also, the standardization makes it possible for these learning histories at particular institution to be stored at a central location commonly referred to as Learning Record Store (LRS). We leverage on the notion of standardization and ownership of LRS by institutions to develop a peer-to-peer network or a blockchain network of institutions' LRSs with privacy and access regulated by smart contracts.

A blockchain is a distributed database of records or public ledger of all transactions or digital events that have been executed and shared among participants [8]. Each transaction or block represent a valid state transition on the network. In a financial scenario for example, state transition could mean a transfer of money from one customer to another. With state transitions being the underlying mechanism for carrying out processes on the blockchain, it is also possible to define instructions that express other real-world processes and use such to determine when and how transactions should be processed on the blockchain. This representation of real-world processes on a blockchain loosely defines smart contracts.

Also, apart from the genesis block, all blocks contain a hash of the previous block. Before a new block is added to the network, some nodes on the network (called miners) offer to add new blocks to the ledger by competing among themselves to solve a computationally intensive puzzle known as the Proof of Work (PoW). The process of solving the PoW puzzle is known as mining. The first node to solve this puzzle is rewarded with the fees associated with the transaction.

Below, we identify some of the limitations of current systems and how the blockchain technology can be used to overcome such limitations.

2. Limitations of Current Approaches

Current alternatives to solving the problem of connecting learning histories across different institutions include the use of Learning Interoperability Tools (LTI), Single Sign On (SSO) and Open Authentications (OAuth). While these technologies solve problems basically in the aspect of authentication, they are still limited in the following areas:

2.1 Connecting Learning Histories

While learners typically move from one provider's learning platform to another, their learning records are stored distinctly and in a disconnected fashion in separate LRSs. Consequently, each system has to pay the cost of growing learner's data from scratch even for very simple cases. While this might not be a repeated effort for first time learners, it is almost impossible to tell if they are truly first timers or not. This also causes a "cold start" problem in training recommender systems due to unavailability of students' previous learning actions [16]. In the current systems design in

連絡先：Ocheja Patrick，京都大学，Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501 JAPAN. Phone: +81-75-753-753, ocheja.patrick@gmail.com

figure 1, we see same users having repeated accounts on different learning systems with no connection between these accounts.

In this work, we propose leveraging on the distributed consensus and immutability features of the blockchain technology to connect learning histories. Nodes on the blockchain would represent institutions with their learning platforms and a single LRS connected to all their platforms. Learning histories will then be broadcast on from these institutions for mining and upon successful mining, it will be added as a ledger entry. Subsequent entries will be chained to previous entries as immutable timestamped blocks. The proposed design is shown in figure 1.

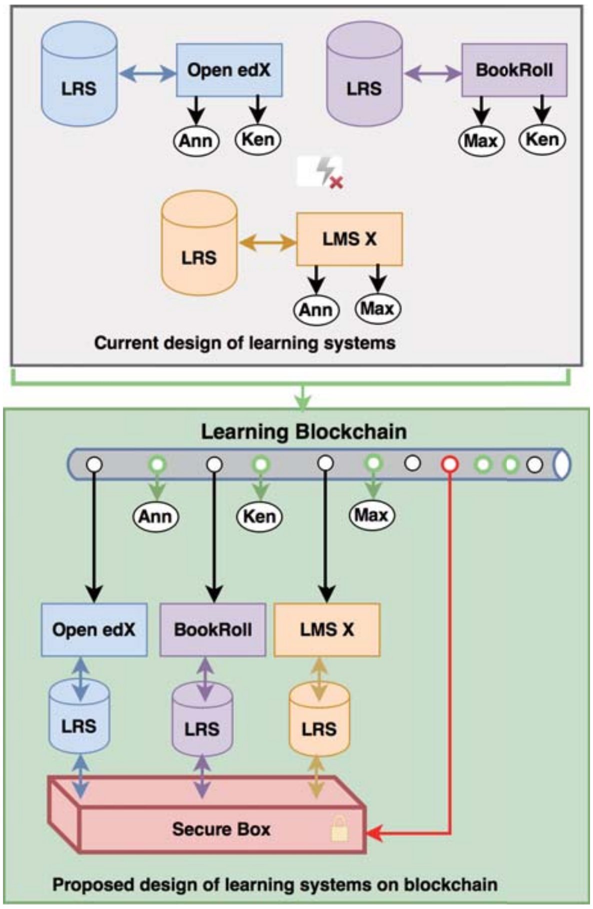


Figure 1. Current learning systems design vs proposed design of learning blockchain.

2.2 Privacy, Security and Access Control

Although, learning analytics helps in improving the performance of learners [3] [4], Alan and Kyle [5] in one wide and four narrow questions about conditions for learner's privacy, argue that whatever the gains of learning analytics are, they must be commensurate to respecting learner's privacy and associated rights. The psychological trauma that could result from a single point of privacy compromise can be quite devastating as it is possible to reveal more confidential information from a single point [6].

Smart contracts on the blockchain can be used to guarantee privacy, security and well audited access control. While policies regarding access of learning histories may vary depending on the content of the learning history, we propose that the different standards for learning histories be observed and learning histories smart contracts should be developed accordingly.

2.3 Integrating Research and Production Systems

Availability of learning data for research fosters innovation. In cases where learning data are collected from production and/or research systems, learning analytics researchers are often faced with the heinous task of anonymizing personally identifying information in order to protect privacy of stakeholders and consequently impacting negatively on personalized results [7]. As real-time learning data becomes more desirable for learning analytics research [7], it is crucial to develop new ideas on how to carry out such seamless integration and interoperability of both research and production systems while maintaining privacy of stakeholders involved.

With the blockchain, research activities can be conducted on real time data with very minimal delay compared to current situation. We have earlier shown how institutions can request access to student's learning histories in [20]. Research systems can coexist on the blockchain with production systems in similar manner.

3. Implementation of a Blockchain of Learning Record Stores (BLRS).

In figure 1, we propose a paradigm shift from current implementations of learning management systems and platforms to the blockchain technology. Block content represent pointers to learning data with ownership and access policies. Nodes on the peer-to-peer network represent learning providers and learners. Learning activities performed by learners on the learning platforms of learning providers on the network are logged on the blockchain as string representation of queries that can be executed on an external database of learning providers to retrieve such activities. To ensure data consistency and immutability, at block creation time, we execute accompanying queries on the external database and include a cryptographic hash of obtained result as part of the block information. Future response from the execution of this query can be compared to the stored hash and if different, the response is invalid and rejected. We propose herein a secure box for executing these queries against providers' databases with reference to the blockchain network in order to maintain established permissions.

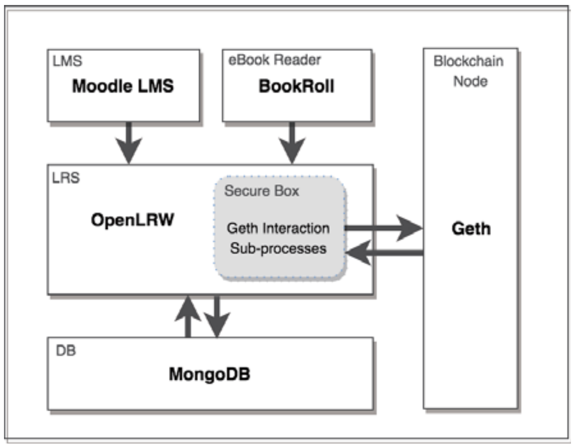


Figure 2. System Architecture – one Institution

3.1 System Architecture

Figure 2 shows a typical setup of our implementation. We use Moodle LMS[18] and BookRoll [7] as the learning platforms. All learning records emitted on these platforms through learning activities of learners are stored in a central database (MongoDB) through OpenLRW [19]. These learning records are either formatted in xAPI [1] or Caliper standard [2]. We also provide an implementation of a subroutine for retrieving records from the MongoDB through the wrappers on OpenLRW and writing them to the blockchain which in this case is the open source Ethereum blockchain implemented in Go programming language (Geth) [17].

3.2 System Access and Privacy Control

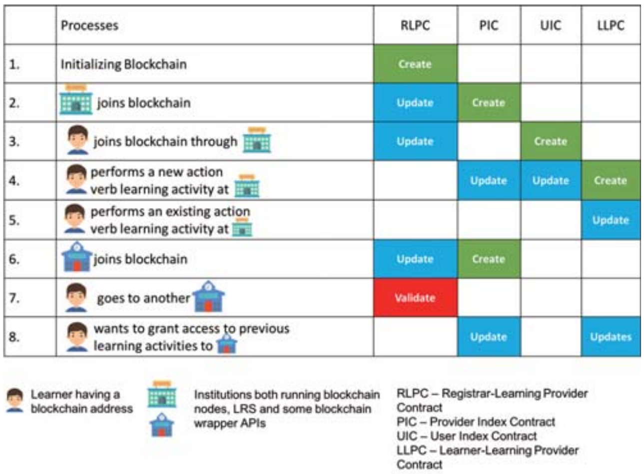
We propose smart contracts that contain learning data access permissions, ownership and a mapping of the two. The state transition functions of these smart contracts can be modified to reflect the conditions that must be met before data read or write access is granted. We define three main smart contracts namely; Registrar – Learning Provider Contract (RLPC), Learner – Learning Provider Contract (LLPC) and Index Contract (IC) for both Providers and Learners.

1. *Registrar – Learning Provider Contract (RLPC).* This contract controls how organizations and institutions become authorized learning providers on the learning blockchain. As these requirements are administratively decided, we propose that typical implementations should consider existing structures for establishing communication and accessing information in institutions and organizations.
2. *Learner – Learning Provider Contract (LLPC).* It represents a proof of existence of a learner's learning data on a learning provider's platform. It contains information about the owning learner, address of learning provider's LRS or database with required authentication parameters, queries that can be executed on learning provider's LRS to retrieve learning data, a hash of expected learning data for ensuring data has not been tampered with and a list of access permissions.
3. *Index Contracts (IC).* An Index Contract contains all LLPCs established between learners and learning providers and by extension, the trail of all learning activities on the blockchain. This is necessary to provide a mechanism for fast lookup of entries and access permissions on the blockchain. We use a hash-table based implementation for the list mapping learners to their LLPCs and another one mapping learning providers to LLPCs they have with learners and with other learning providers that learners have granted access.

3.3 Processes Involved in BLRS

To setup a BLRS, it required to have at least one institution with the setup in Figure 2. The RLPC is then installed on the blockchain node. With this, all institutions that wish to join the blockchain will have to request to be registered by having a similar setup as in Figure 2 and then sending a registration request to the RLPC which was initially installed on the hosting institution's blockchain node. Upon approval, the RLPC is update with their information and a PIC is created for them.

Learners that opt to have their learning records on the blockchain will have to go through the account setup process. This process handles the generation of blockchain address for the learner, creation of an Index Contract – UIC and the final phase of registering the generated blockchain address and UIC address in the RLPC.



	Processes	RLPC	PIC	UIC	LLPC
1.	Initializing Blockchain	Create			
2.	joins blockchain	Update	Create		
3.	joins blockchain through	Update		Create	
4.	performs a new action verb learning activity at		Update	Update	Create
5.	performs an existing action verb learning activity at				Update
6.	joins blockchain	Update	Create		
7.	goes to another	Validate			
8.	wants to grant access to previous learning activities to		Update		Updates

Figure 3. Processes involved in BLRS.

On the blockchain, learning records are uniquely grouped using the action verb field and the user's blockchain address and writing learning histories entails performing at least one transaction on the blockchain. The process begins with retrieving the action verb of the learning record and converting it to a corresponding hexadecimal number. This is required because we want to optimize gas usage on the blockchain; writing strings of variable length require more computational resources in solving the Proof of Work especially when the string is lengthy. After converting the action verb to hexadecimal equivalent, we then query the blockchain to know if a smart contract based on this action verb exists for this user. If it does, we retrieve the smart contract and simply update it with the current learning record's query string and query result hash. If no such smart contract exists for this action verb, we create the smart contract and update the index contracts of both the provider and the learner. The latter case will require four transactions which must be mined on the blockchain.

4. On-going Evaluation

To understand the resource requirements and comparative advantages of connecting learning histories of students using blockchain technology, we have setup test systems following the architecture in figure 2. We have also observed resource requirements in terms of having a single node (or institution) processing all transactions for all the students in their school with no other institution being part of the network. Currently, we are evaluating more scenarios of multiple institutions acting as miners. Our evaluations include integration tests on installation, implementation and interaction with the different smart contracts. We hope to present a side-by-side comparison with the previously identified alternative systems currently being used.

From this implementation, we observe that while some of the transactions on the blockchain require very minimal resources (such as the blockchain address issuing transaction), others require some amount of time; typically, about 2 minutes. Also, we observed that the more the available mining nodes on the network, the faster the transactions are processed (test results to be provided).

We hope to provide more information about the results from our on-going evaluation in our presentation at the OS-9 session.

5. Discussion

For an on-demand connection of learning records, this time might be acceptable. But for real-time connection of learning records it might pose a challenge. Our observation of transaction speed being proportional to mining nodes available ideally follows the tenets of a decentralized network where the best throughput is achieved if everyone mined their own transactions. While it might be difficult to achieve a system where all learners mine their own learning records, it will be interesting to consider alternative approaches to improving on transaction processing time by leveraging on client-side-browser-based mining nodes.

However, one very important concern is the sustainability of the blockchain technology due to the large computing and energy resource requirements. We are aware of this limitation and hence, we have constrained current implementation to use the institution's central resource.

6. Conclusion and Future Work

In this work, we introduced the concept of connecting learning records using the blockchain. We provided some core aspects of our current implementation which is still been developed. An important aspect to be considered is defining and enforcing existing user data privacy policies on the learning records using smart contracts. While our implementation considers very top-level approach of representing these permissions, it will be necessary to understand the implications of having 'action verb-based' privacy definitions. Also, we also propose that a further research should be done on how learners can write their own smart contracts using familiar concepts and enforce them on their learning records.

In our future work, we will provide more concrete results on resource requirement, throughput, and a close comparison to alternative systems currently been used.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 16H06304.

References

[1] Advanced Distributed Learning. (2016). Experience API (xAPI) Specification.

[2] IMS Global Learning Consortium. (2015). Caliper Analytics.

[3] Fumiya Okubo, Takayoshi Yamashita, Atsushi Shimada, Hiroaki Ogata, A Neural Network Approach for Students' Performance Prediction, LAK 2017, pp.598-599, 2017.3.

[4] Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education. JISC.

[5] Alan Rubel and Kyle M. L. Jones. 2016. Student privacy in learning analytics: An information ethics perspective. The Information Society. Vol. 32(2). 143-159. DOI: http://dx.doi.org/10.1080/01972243.2016.1130502

[6] Omer Tene and Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 11 Nw. J. Tech. & Intell. Prop. 239 (2013).

[7] Brendan Flanagan and Hiroaki Ogata. 2017. Integration of Learning Analytics Research and Production Systems While Protecting Privacy. Chen, W. et al. (Eds.) (2017). *Proceedings of the 25th International Conference on Computers in Education.* New Zealand: Asia Pacific Society for Computers in Education. (in press)

[8] M. Crosby et al. 2015. Blockchain Technology; Beyond Bitcoin, Sutardja Center for Entrepreneurship & Technology. Berkeley Engineering. Retrieved from http://scet.berkeley.edu/wp-content/uploads/BlockchainPaper.pdf

[9] S. Nakamoto. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. White Paper.

[10] A. Azaria, A. Ekblaw et al., MedRec: Using blockchain for medical data access and permission management, In 2016 2nd International Conference on Open and Big Data (OBD). Institute of Electrical and Electronics Engineers (IEEE), Aug. 2016.

[11] H. Kalodner et al. An empirical study of Namecoin and lessons for decentralized namespace design. WEIS '15: Proceedings of the 14th Workshop on the Economics of Information Security, June 2015.

[12] G. Zyskind et al. Decentralizing privacy: Using blockchain to protect personal data. Security and Privacy Workshops (SPW) 2015 IEEE. IEEE pp. 180-184 2015.

[13] Sony Global Education. 2017. Sony Develops System for Authentication, Sharing, and Rights Management Blockchain Technology. News Release. Retrieved from https://www.sony.net/SonyInfo/News/Press/201708/17-071E/index.html

[14] Buterin, V. 2013. Ethereum White Paper. Retrieved from https://github.com/ethereum/wiki/wiki/White-Paper

[15] SHA-2 Standard. Secure Hash Standard FIPS PUB 180-4. Retrieved from https://csrc.nist.gov/publications/detail/fips/180/4/final

[16] Barnes, T., and Stamper, J. 2008. Toward automatic hint generation for logic proof tutoring using historical student data. In *International Conference on Intelligent Tutoring Systems* (pp. 373-382). Springer, Berlin, Heidelberg.

[17] Ethereum in Go Language. https://github.com/ethereum/go-ethereum

[18] Moodle Learning Management System. https://github.com/moodle/moodle

[19] Open Learning Record Warehouse. https://github.com/Apereo-Learning-Analytics-Initiative/OpenLRW

[20] Patrick Ocheja, Brendan Flanagan, and Hiroaki Ogata. 2018. Connecting decentralized learning records: a blockchain based learning analytics platform. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. ACM, New York, NY, USA, 265-269. DOI: https://doi.org/10.1145/3170358.3170365