Connecting Decentralized Learning Records: A Blockchain Based Learning Analytics Platform

Patrick Ocheja
Graduate School of Informatics, Kyoto University
Kyoto, Japan
ocheja.ileanwa.65s@st.kyoto-u.ac.jp

Brendan Flanagan
Academic Center for Computing and Media Studies, Kyoto University
Kyoto, Japan
flanagan.brendanjohn.4n@kyoto-u.ac.jp

Hiroaki Ogata
Academic Center for Computing and Media Studies, Kyoto University
Kyoto, Japan
hiroaki.ogata@gmail.com

ABSTRACT

As Learners move from one learning environment to another, there is a key necessity of taking with them a proof of previous learning achievements or experiences. In most cases, this is either expressed in terms of receipt of scores or a certificate of completion. While this may be sufficient for enrollment and other administrative decisions, it poses some limitations to the depth of learning analytics and consequently a slow onboarding process. Also, with different institutions having their learning data isolated from each other, it becomes more difficult to easily access a learner's learning history for all learning activities on other systems. In this paper, we propose a blockchain based approach for connecting learning data across different Learning Management Systems (LMS), Learning Record Stores (LRS), institutions and organizations. Leveraging on unique properties of blockchain technology, we also propose solutions to ensuring learning data consistency, availability, immutability, security, privacy and access control.

CCS CONCEPTS

• Applied computing~Education~Computer-managed instruction • Applied computing~Education~Learning management systems • Security and privacy~Security services~Privacy-preserving protocols

KEYWORDS

Learning analytics; blockchain; learning data; smart contracts; learning management systems; learning record store; privacy

ACM Reference Format:

P. Ocheja, B. Flanagan, and H. Ogata. 2018. Connecting Decentralized Learning Records: A Blockchain Based Learning Analytics Platform. In *LAK'18: Proceedings of International Conference on Learning Analytics and Knowledge, Sydney, Australia, March March 7–9, 2018, Sydney, NSW, Australia. ACM, New York, NY, USA, 5 pages.*
https://doi.org/10.1145/3170358.3170365

LAK'18, March 7–9, 2018, Sydney, NSW, Australia

ACM ISBN 978-1-4503-6400-3/18/03…$15.00
https://doi.org/10.1145/3170358.3170365

1 INTRODUCTION

Learning data reflect the activities performed by learners while learning. From information on a learner's behavior to performance in quizzes and assignments, these data form a reference point for evaluating and improving engagement and performance towards realization of learning goals. With many learning organizations and institutions, the multiplicity of different implementations of learning platforms is inevitable. As such, it becomes necessary to ensure a standard for learning data. Common standards such as Tin Can Experience API [1], IMS Caliper Discovery API [2] have been developed to help reduce the burden of system interoperability. It is on the awareness of these standards that learning data silos otherwise known as Learning Record Stores (LRS) are maintained. These record stores form the backbone for learning analytics.

1.1 Limitations of Learning Analytics Platforms

Despite the availability of reference standards for maintaining learning data on an LRS, it is still difficult to achieve interoperability without some limitations. These problems include:

- Connecting learning histories of a learner on different learning platforms on a single immutable trail.
- Ensuring privacy of learners' records with ease of access control.
- Integrating research and production systems for advancing learning.

1.1.1 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]. Proposed systems should allow learners to take their learning data with them in the same way they can take their certificates easily from one institution to another.

1.1.2 Privacy, Security and Access Control. This is another challenge faced when sharing learning records with third parties. 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]. Proposed systems should ensure prioritization of learner's privacy and learners should be in control of their learning data.

1.1.3 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.

1.2 Blockchain Features as a Solution

This work addresses previously identified limitations of current systems in enhancing learning analytics. We propose solutions to mobility of learner's learning records, distributed consensus in maintaining learning history, privacy and access control mechanisms with prioritized learner's interest and interoperability of different systems (production and research). 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]. Below, we identify some of the features of blockchain technology that are key to our proposed solution.

1.3.1 Distributed Consensus and Immutability Features. With its first implementation in Bitcoin [9], blockchain technology is based on a distributed consensus where nodes on the network have access to and keep track of all events that occur on the network. Ledger entries are stored as timestamped, chained immutable blocks. To ensure security and consistency of ledger entries, some nodes on the network offer to add new blocks to the ledger by competing among themselves to solve a computationally intensive puzzle known as the Proof of Work. These nodes are called miners and are rewarded for being the first to provide a correct solution to the Proof of Work. The computing power required for solving this puzzle makes it more difficult to rewrite blocks as such rewrite by dishonest nodes would require resolving associated Proof of Work and acceptance of such solution by honest nodes. These features of blockchain technology provide answers to connecting different learning records from different learning providers with high data consistency.

1.3.2 Smart Contract-based Privacy, Security and Access Control. The blockchain technology has a smart contract feature that facilitates enforcing the terms of agreement between two parties in a contract; in this case, between learners and learning providers or between learning providers. We propose policies that are deployable on the blockchain to control data access and ensure privacy of learner's records and mutual interests of learning providers.

1.3.3 Single Ledger, Multiple Participants. We leverage on the distributed consensus and single ledger-multiple-participants features of the blockchain technology to enhance interoperability of both research and production systems. We propose Learning Blockchain APIs and Datastore Wrappers for ensuring seamless and secure communications between the blockchain and LRSs of learning providers. We suggest potential candidates for enforcing non-intrusive access request and provision for foreign systems.

1.3 Related Work

We are aware of only one other effort in the application of blockchain technology to education; Sony [13]. Apart from a press release [13], no specific methods or summary of technical approach has been published yet. To the best of our knowledge, we are the first to provide a system design for a blockchain based network of learning records for learning analytics.

In fields other than education and learning analytics, there is existing research on applying blockchain technology to non-financial products, such as: medical information [10] and domain name registry [11]. Zyskind et al's work on using blockchain to protect personal data provides insight on achieving privacy preservation on a decentralized network with user control and auditing [12]. While these ideas are fundamental to our discovery of our novel approach, there are many aspects of learning systems that present unique problems that need to be solved, such as: connecting distributed or disconnected learning data, smart contract based privacy and access control frameworks, and interoperability of different learning systems for both research and production environments. This paper proposes an innovative blockchain based system with important modifications to address the specific needs of education and learning systems.

2 PROPOSED BLOCKCHAIN FOR LEARNING ANALYTICS

2.1 Overview

Our design will specify processes for creating, adding and retrieving learning data on the blockchain. 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.

In the next sections, we will discuss further the design of our proposed system and the underlying principles.

2.2 Ethereum Blockchain

Technically, a blockchain can be viewed as consisting of state transition machines. In this case, a state transition machine's state is identified by ownership status of transaction outcomes (as in bitcoin) and a state transition function that specify conditions for valid state transition. To create a new state, an existing state with valid transaction is passed as input to the state transition function which outputs a new state on the blockchain. Thus, the blocks logically represent all the valid state transitions on the network. One may think that with such an easy sequence of steps, states could be easily generated uncontrollably or even modified. However, this is not true because of the existence of the Proof of Work. The Proof of Work required before new states are created makes it computationally difficult to modify states and controls the rate at which new states are generated. For example, if the Proof of Work requires that the double-SHA256 hash [15] of every state, treated as a 256-bit number, must be less than 2^{180}, it will require a successful node on the network to make an average of approximately 2^{76} tries before a valid state is found. When a valid state is found, the information is broadcast to all nodes on the network to verify correctness of the solution before the new state is accepted and the successful node is rewarded. Subsequently, nodes can query the state machines at any time to obtain a correct and valid state already verified by everyone else.

Thus, it is possible to express real-world processes as states and state transition functions. This code representation of real-world processes on a blockchain loosely defines smart contracts. Although present in bitcoin blockchain, Ethereum (eth) is the first to implement a blockchain with a Turing-complete smart contract programming feature [14]. Being Turing-complete is important because it enables writing programs (especially with loop directives) in fewer instructions with efficient use of space. The concept of smart contract lies at the heart of our proposed design as it makes it feasible to enforce required policies and processes by expressing them as executable codes on the blockchain.

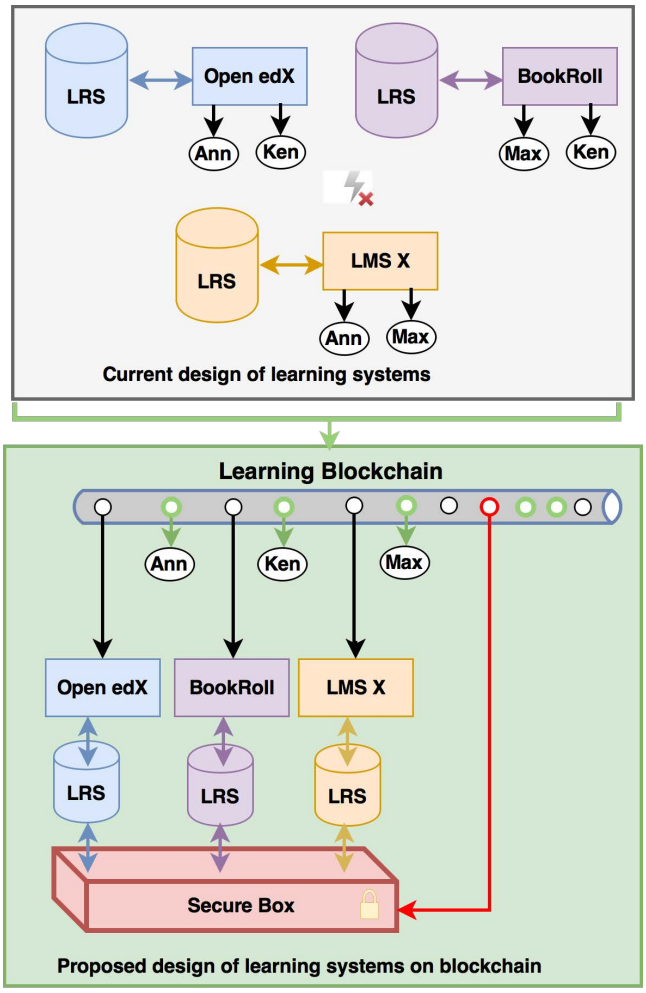

Figure 1. Current learning systems design vs proposed design of learning blockchain

2.3 System Access and Privacy Control

We propose contracts that contain learning data access permissions, ownership and a mapping of the two. The state transition functions of these contracts can be modified to reflect the conditions that must be met before data read or write access is granted. In figure 2, we show the structure of the three main smart contracts namely; Registrar – Learning Provider Contract (RLPC), Learner – Learning Provider Contract (LLPC) and Index Contract (IC).

2.3.1 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. An example could be the use of special identifiers (ID-1, ID-2, and ID-3 in figure 2) and/or tokens to verify that a node requesting access to the network is actually

a known party to the other nodes. This and other conditions can be coded into the RLPC.

2.3.2 Learner – Learning Provider Contract (LLPC). It represents a proof of existence of a learner's learning data on a learning provider's platform.

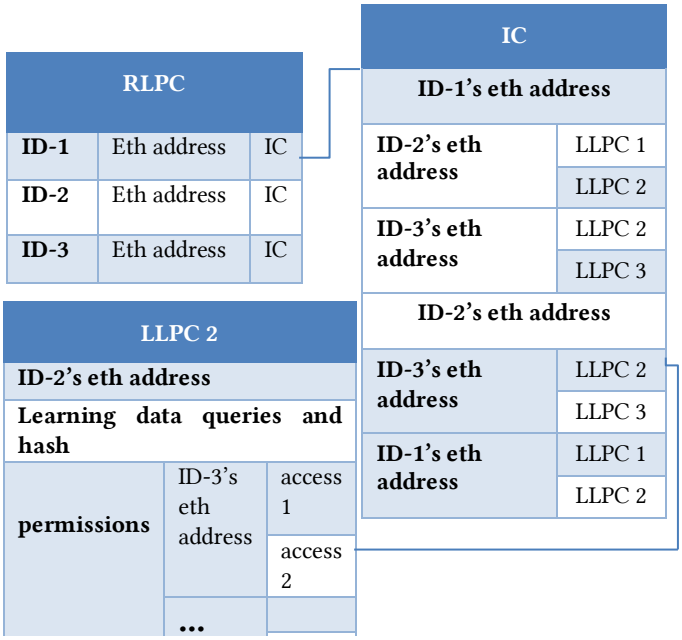

Figure 2. Proposed smart contracts on the learning blockchain.

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. LLP Contract empower learners with the ability of controlling who can view their learning data by maintaining a list of access permissions granted to other learning providers.

2.3.3 Index Contract (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 suggest 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 or learning providers that learners have granted access permission.

For example, in figure 2, we show two entries of the Index Contract. ID-2's entry refers to a mapping for the learner to LLPCs while ID-1 and ID-3 entries represent a learning provider to LLPCs mapping. As with all other smart contracts, the IC is broadcast on the network. This makes it possible for learners and learning providers to leave and return to the network without losing their data. In a case where a learning provider wants to request access to learner's learning data on another provider's platform, a disabled IC is issued and only becomes active when the learner approves the request. It is important to also note that even though a node does not have permission to a LLPC which means no access to any learner's learning data, it still maintains a reference (a stewardship requirement) to the LLPCs issued and/or accessible by other participants on the network.

2.4 System Nodes and Utilities

In our design, we require learning providers to join the blockchain through a node managed by them. It is not required for learners to maintain a node, they can easily join the learning blockchain by registering on any learning provider of their choice. The learning provider in turn creates an account on the blockchain for the learner and issue them a public-private key pair for tracking their content on the blockchain from any learning provider's platform that is on the blockchain. We propose additional software components for managing data on the blockchain. This include Learning Blockchain APIs (LB API), a Secure Box with LRS Wrappers and customized Ethereum Client.

2.4.1 Learning Blockchain APIs (LB APIs). With the understanding that interacting with the blockchain might be a tough hurdle especially for programmers that are not familiar with blockchain technology, we propose development of APIs for communicating with the learning blockchain. This should abstract processes such as creating transactions, monitoring its success or failure, creating and accessing smart contracts and triggering mining activities. However, our design assumes that nodes on the network have the required resources (gas measured in ethers; a unit representing the worth of computing time) for processing transactions. Future works will provide incentive mechanisms for mining.

2.4.2 Secure Box with LRS Wrappers. We propose a tool for keeping all interactions between the learning blockchain and LRS of different learning providers secure. We establish all of such communications within a Secure Box that is bundled with all nodes. Within this Secure Box, Database Wrappers are provided to take care of the differences that exist between LRS of learning providers. We also keep a connection to the blockchain to verify that all query execution request made through the Secure Box are coming from authorized Learners or Learning Providers specified on the LLPC from which the query was obtained.

2.4.3 Ethereum Client. This is a customized version of Ethereum blockchain network. It has full features required of a blockchain network with peer-to-peer networking, contracts, transaction handling and mining capabilities. In this work and future works, we build on Go Ethereum. Our proposed modifications include adding services to actively monitor contracts creation and index them appropriately on the Index Contract. The client should also be aware of unique identifier mappings to eth addresses as contained in RLPC.

In figure 3, we show sequence of activities that occur on the blockchain and how they are handled. At S_0, the blockchain contains only the boot node, RLPC and a Secure Box. KU node then attempts to join the network which prompts verification with established rules in RLPC. Upon successful verification, KU

is added as a valid participant and an IC is generated. Learner A (L-A) visits KU's platform and since it is its first visit to any node on the network, a new account is created for L-A at S_2. Subsequent learning activities leading to generation of learning data are logged on the blockchain as LLPC-A_n.

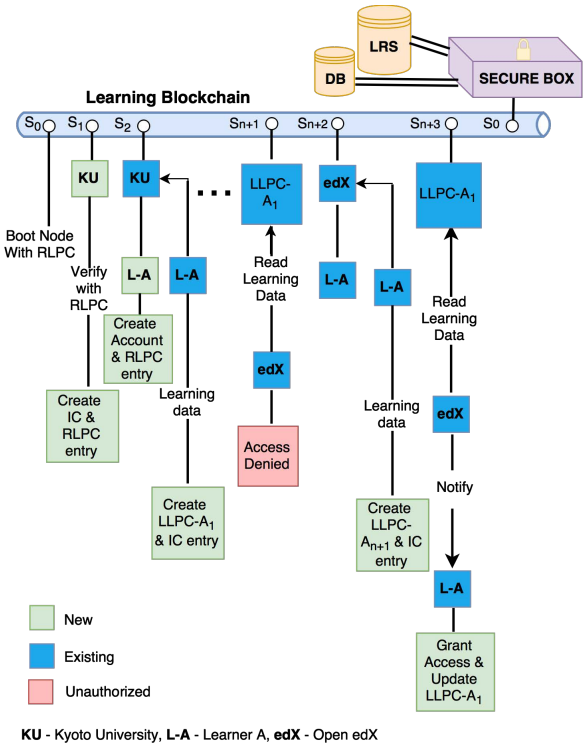

Figure 3. Sample process of registering and accessing blockchain information.

At S_{n+1}, edX attempts to read the learning data (LLPC-A_1) of L-A, this is outrightly rejected as there is no proof of edX being aware of the existence of L-A. Later on, L-A decides to visit edX platform and provides their blockchain information to edX. Now, edX knows of the existence of L-A. This means that further request to access L-A's learning data will be forwarded to L-A for approval. If approved, the permission is written on the LLPC and access to the learning data is granted. The queries stored on the LLPCs are then executed by the Secure Box on the LRS or DB.

3 CONCLUSION AND FUTURE WORK

We have shown how the blockchain technology can be used to establish connections between decentralized learning systems and maintaining a continuous log of learning activities performed by learners. Leveraging on smart contracts, we have also proposed how privacy and security policies can be implemented on the platform. In future works, we will build on this and provide implementations. We will also explore possibilities of storing actual learning data on the blockchain with a good compromise on storage requirements.

ACKNOWLEDGMENT

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

REFERENCES

[1] Advanced Distributed Learning. (2016). Experience API (xAPI) Specification. Retrieved from http://github.com/adlnet/xAPI-Spec/
[2] IMS Global Learning Consortium. (2015). Caliper Analytics. Retrieved from http://www.imsglobal.org/activity/caliper
[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. Retrieved from http://repository.jisc.ac.uk/6560/1/learning-analytics_and_student_success.pdf
[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). Retrieved from http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1
[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.