Preserving the Privacy for University Data Using Blockchain and Attribute-based Encryption

Soe Myint Myat
Computer Science Department
Myanmar Aerospace Engineering University
Meiktila, Myanmar
soemyintmyat@ucsy.edu.mm

Than Naing Soe
University of Computer Studies, Myitkyina
Myanmar
kothannaingsoe@gmail.com

Abstract

An effective IT solution is required for the administrative processes which is the core of university. The tamper resistance, transparency and auditability for university management data are very important to avoid the corruption. Blockchain occupies the immutability and irreversibility properties thus, it becomes a potential solution. Regrettably, there are some challenges such as privacy concern and limited storage exists in blockchain technology. In this work, a blockchain-based university data storage model is proposed and such barriers of blockchain are handled. The proposed model achieves the tamper resistance, transparency and auditability feature by using blockchain technology. CP-ABE and other cryptographic techniques are used to provide the fine-grained access and privacy preserving facilities. The security analysis is performed and shows that our approach is tamper resistant and provably secure for privacy.

Keywords—*data security, tamper resistance, blockchain, privacy, fine-grained access*

I. INTRODUCTION

In the premise of industrial 4.0 revolutions [1], human society will be stimulated with the higher education which is an essential factor for human development. Thus, a large number of information systems such as office automation system, teaching system, administrative system, personnel system and asset system are built in universities to improve the efficiency of staffs [1]. This work will make dialectical, complicated an interesting prospects of human centric characteristic. The research, services and teaching will be changed to different ways and new forms of universities will be emerged in the fourth industrial revolution. Thus, the tamper resistance, transparency and auditability for university management data are very important to avoid the corruption. Then again, the industrial 4.0 also takes the popular technology, blockchain which occupies the transparency, immutability and cryptographic verifiability properties. Thus, blockchain become a candidate solution for university data.

Contrary to the expectation, the blockchain technology also contains some obstacles in using for university data. One of the obvious requirements for university data is fine grained access that allows the only authorized user to access the certain data. For instance, if the university data of all departments are transparently stored on blockchain, any faculty or everybody who can access the blockchain can access any document. However, the blockchain stores cryptographically verified data; it does not encrypt the data at all [2]. In some private blockchain system like Hyperledger Fabric [3], the participation in network can be regulated, however, the university data still needs to allow only certain users to access a specific data. Consequently, the privacy issue has to be handled. The append-only property of blockchain becomes a barrier for a user revocation feature which allows eliminating permission of the access on the university data for specified individuals. The explosive growth of the university data causes the available issue for the limited blockchain storage. Thus, how to use blockchain technology for university data as an underlying mechanism is still an issue.

To provide the confidentiality and to preserve the privacy, encryption is a promising way. However, there exist some traditional technology such as password based and classical public key encryption based approaches, the individual user needs to maintain too much secret information (decryption keys or passwords) for accessing the multiple files. The attribute-based encryption (ABE) [4] uses the user attribute set as the public key instead of using random string as the public key. The further development of ABE encrypts the message with an access policy and it is called Ciphertext Policy Attribute-based Encryption (CP-ABE) [5]. In CP-ABE scheme, only the user who

has the attributes that meet the access policy requirement can decrypt the encrypted message thus, CP-ABE can support more efficient access control mechanism.

In this work, the blockchain technology, cloud storage and CP-ABE encryption technique are used to preserve the privacy for the university data. The blockchain technology is used to support the tamper resistance, auditability and transparency of university data. The cloud storage is used to overcome the availability issue of blockchain and CP-ABE is used to support the fine grained access on university data. Thus, our approach requires only one secret key per user while other systems such as password based systems require multiple passwords for multiple files. Similarly, our approach requires encrypting the data only one time while public key crypto systems require multiple times for encryption for a document for multiple users. At the same time, our approach can support tamper resistant, auditability and transparent property for university management data.

The followings are the rest of the paper. Section 2 discusses about some existing related works. Section 3 discusses some technology and knowledge that support in developing our system such as blockchain technology, CP-ABE encryption and other cryptographic primitives. The detailed of our proposed approach is discussed in section 4 and security analysis is performed in section 5. Finally, the paper is concluded in section 6.

II. RELATED WORKS

Blockchain is used as the storage for diverse items such as academic work, attendance, certificates and awarding of a university degree in the education purpose [6], [7]. The distributed and transparent properties of blockchain are used to reduce the fraud in academic items such as certificates and informal context such as misrepresentation of knowledge, background and skill [8]. Most of these existing works are trying to use some properties of blockchain such as transparency and there are some challenges such as confidentiality, privacy and availability for using blockchain.

In [9] the fine grained access control for encrypted data is firstly developed with a variant of ABE. However, this approach cannot support complex policies require for university data. To support such sophisticated policies, CP-ABE which describe the users with various attributes is more suitable. The CP-ABE algorithm is firstly introduced by Bethencourt et al. in [6]. They used access tree in encrypting the message and "Lagrange Interpolation" is used in decryption. Then, a comparative attribute-based encryption is proposed in [10] and the proposed method is illustrated with an example of telemedicine. Moreover, an implementation of Functional Encryption is proposed in [11]. Most of the existing works tried to use attribute-based encryption for healthcare data.

There is no universally accepted standard definition for the term "privacy". Privacy encompasses with several concepts such as anonymity, pseudonymi ty, unlinkability, unobservability and revocability of consent. The blockchain support most of these concepts, thus, we try to use the benefits of blockchain technology for university data and the cloud technology and CP-ABE technology is used to handle some blockchain issue for using in university data storage.

III. PRELIMINARIES

A. Blockchain

The blockchain technology is introduced by Satoshi Nakamoto with the crypto-currency called Bitcoin [2]. Actually, the blockchain is a distributed database and each data storage structure called block links each other as a chain [12]. A block is a special storage structure of blockchain and it maintains the hash value of the previous block to form a chain. By forming a chain, the block is immutable to modification [13]. The block also maintains other items such as payload, timestamp and signature of the contributor. Payload of the block may vary according to the various applications. The payload can be any asset or item such as the content of the transaction, an address pointer of the original data or some other information. The timestamp is used to order the blocks in chronological fashion. The signature of contributor shows the generator of the block. Generally, blockchain network includes two main entities, miners who produce new blocks and verifiers who verify the new blocks. In generating the new block, consensus mechanism is usually used. In this work the metadata will be stored on blockchain as verified data.

There are three categories of blockchain in general. They are permission-less blockchain also called public blockchain, permission blockchain also called private blockchain and consortium blockchain. The permission-less blockchain is public in nature and everyone can participate in the blockchain system. Bitcoin is a good example for permission-less blockchain system. In the group of permission blockchains, the access right and participation on

blockchain network is controlled by an organization in private blockchain while several organizations manage the consortium blockchain. The permission blockchain (private blockchain or consortium blockchain) is suitable to store and manage the university data according to the architecture of our work.

B. Ciphertext-policy Attribute-based Encryption

The ciphertext-policy attribute-based encryption (CP-ABE) [5] is a one-to-many encryption scheme. CP-ABE allows the multiple users to access the encrypted data. CP-ABE use a set of attribute in identifying the user with the decrypting key called CP-ABE private key. By using a list of attribute which correspond to the authorized users, the data owner can specify various access policy. The policy is then embedded into the encrypted data. For example, the policy can be expressed as follow.

Policy P = "(Rector) OR (Head_of_Department) AND (Department_1)"

If the user processes the attributes 'Rector' or 'Head of Department from Department 1', the ciphertext can be decrypted. Otherwise, the data cannot be decrypted. Generally, there are four main steps in CP-ABE scheme. In the first step, a master secret key MSK and a public parameter PK are generated and this step is called setup phase. The public parameter PK contains the generator g, g^{β}, and an efficiently computable symmetric bilinear mape $(g, g)^{\alpha}$. The master secret key MSK contains the value β and g^{α}. The PK can be reveal publicly, and the MSK must be kept secret.

In second phase, the encrypting process is performed. The cipher text CT is output from the set of input which includes plaintext message M, public parameters PK and an access policy T in encryption phase. A set of Boolean formulas is used to create an access policy tree. The figure 1 illustrates the creating the access policy tree from the policy P.

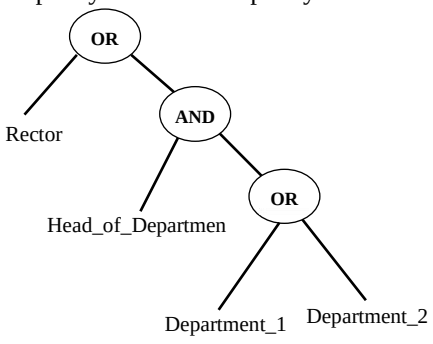

Figure 1. Sample access policy tree

The third step is the key-generation phase. The private key SK which associate to the set of user attributes set S is generated in this phase. The master secret key MSK, the public parameters PK and a set of user attributes S are taken by this process as input. That is why, the attributes are mathematically incorporated into the key.

The fourth step is the decryption phase. The ciphertext CT will be decrypted in the decryption phase, if and only if the access policy tree T is satisfied by theset of attributes associated with the private key SK.

C. Access Structure

Let a set of $n \in N^{+}$ be a set of participants $\{P_1, P_2, \dots, P_n\}$ and a collection $A \subseteq 2^{\{P_1, P_2, \dots, P_n\}}$ is monotone for $\forall B$ and C, and if $B \in A$, $B \subseteq C$, then $C \in A$. Then, an access structure is a collection A of non-empty subsets of $\{P_1, P_2, \dots, P_n\}$, i.e, $A \in 2^{\{P_1, P_2, \cdots, P_n\}} \setminus \{\emptyset\}$ [5].

The sets which are included in A are called as the authorized sets. However, the sets which are not included in A are called unauthorized sets. Attributes take the role of parties in our context. Thus, the access structure A will contain the authorized sets of attributes.

D. Bilinear Map

Let G and $G_T$as two cyclic groups of order p for some large primep. G is a group of points on an elliptic curve over F_P (namely the finite field mod p) is G and a subgroup of a finite field $F_{p^2}^{*}$ (namely the finite field mod p^2) is G_T.

Then, a map $e : G \times G \rightarrow G_T$ is called to be a bilinear map if it satisfies the following properties.

1) Bilinear ($(aP, bQ) = e(P, Q)^{ab}$ for all $P, Q \in G$ and $a, b \in Z_p$).
2) Non-Degenerate (the map does not send all pairs in $G \times G$ to the identity in G_T and observe that since G and G_T are groups of prime order, this implies that if P is a generator of G, then $e(P, P)$ is a generator of G_T)
3) Computable (there is an efficient algorithm to compute $e(P, Q)$ for any $P, Q \in G$)

IV. METHODOLOGY

This section provides the problem scenario of our work and the detailed description about our proposed model.

A. *Problem Scenario*

Generally, in a university, the role hierarchy of the employees may be in the form which is illustrated in Figure 2. If the data are stored in the same location or storage, everyone in the university (all faculties or departments or all participants) can access these data. In reality, there may be various attributes of employees in university environment; however, the Table 1 illustrates some sample attribute list for employees in university.

To support data privacy (authorized access), data protection mechanism which allows only the authorized employees to access the data is required. The authorized employees must occupy the attributes values which can satisfy the access conditions. Thus, the attribute values which are shown in Table 1 are used to generate the decryption keys (private keys) for each group of users.

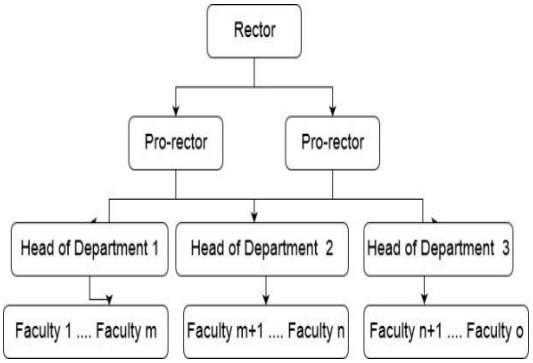

Figure 2. Sample hierarchy of employees in university

For instance, an examination data, S2_D3_ B5_Sem2_Lab_Marks_2016.doc is stored in the system. In this data the subject is S2, the department is D3, and the batch is B5, and the semester is the second semester Sem2. The employees associated with this batch should access it is confidential by this marks data.

Everyone can access this data if the data is stored in common storage. Password-based usage has operational difficulties such as maintaining many passwords although some password protection can be employed for protection. Thus, a new approach is required to store the data in the manner that the data is encrypted; the data is tamper resistant and the data must be accessible by only authorized users.

TABLE I. SAMPLE ATTRIBUTE LIST OF UNIVERSITY EMPLOYEES

Faculty	F1	F2	F3	F4	F5	F6	F7	F8
Department	D1	D1	D3	D1	D2	D1	D1	D1
Designation	Associate Professor	Assistant Professor	Assistant Professor	Professor	Professor	Associate Professor	Assistant Professor	Professor
Role	Lecture	Tutorial	Lab	Coordinator Theory	Lecture	In-charge D1 2018	In-charge D1 2Sem 2019	In-charge D1 2Sem 2018
ID	'ID = 1764'	'ID = 1760'	'ID = 1543'	'ID = 1729'	'ID = 1271'	'ID = 1270'	'ID = 1290'	'ID = 1250'
Batch	B1	B6	B5	B3	B2	B7	B4	B5
Subject	S1	S1	S2	S1	S3	S4	S5	S3
Semester	Sem2	Sem2	Sem2	Sem2	Sem1	Sem2	Sem1	Sem2
Year	2016	2017	2016	2017	2018	2018	2019	2018

B. Proposed method

Attribute-based encryption technique and blockchain propose a privacy preserving so as to support the data owner with the fined grained access control and tamper resistant storage. CP-ABE algorithm encrypts the actual data to support confidentiality. The encrypted data is stored on the cloud storage which can guarantee the availability. The metadata which represents the university data is permanently stored on the blockchain to provide a search and to obtain the tamper resistance property. Figure 3 represents the overall architecture.

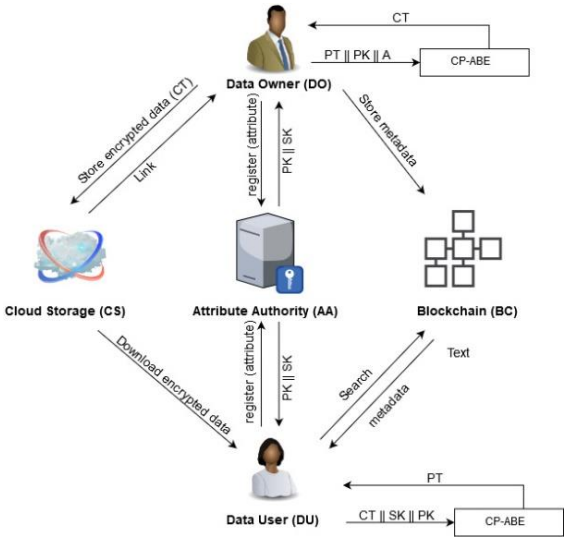

Figure 3. Architecture of the proposed model

There are five main entities that include in proposed model. The first entity is the data owner (DO). DO will access or store the university data. DO has full right to control the over the data, thus, DO will define the access policy to allow or disallow some accesses on data. The second entity is the data user (DU) and DU can access the data with the permission of corresponding DO. Typically, DU can be an entity from several sections. DU can search the information about data on blockchain via the metadata. Then, DU can download the encrypted data from cloud storage (CS). The third entity is the attribute authority (AA) and it is responsible for verifying the attribute of each user and issuing the corresponding key. The other entities are cloud storage (CS) which stores the encrypted data and blockchain (BC) which stores the metadata.

1) Setting the system

All participants in the system must register with the attribute authority to setup the system and perform the initial agreement to support the operations in the model. To start the agreement, AA performs the setup algorithm. The setup algorithm uses the implicit security parameter as an input and it produces a master key MK and the public parameter PK. Then, AA performs the key generation algorithm for each user with their attributes set S and master key MK. At the end of setting the system, AA distributes an appropriate private key SK to corresponding user.

2) Storing Data

The data owner (DO) starts the store operation by performing the encrypt algorithm with an access structure A over the universe of attributes and the sensitive information or message M. Therefore, users who occupy the required attributes set which can satisfy the access structure can decrypt the resulted ciphertext CT. On the other hand, the the access policy which is embedded in the ciphertext CT or CT implicitly maintains the authorized attributes A. Then, the hash code is generated from CT for integrity checking. CT is stored on cloud storage and the link to CT is received. The DO extracts the metadata and includes the hash code and link for search purpose within the system. Finally, DO stores the metadata and other required information about the encrypted data on the blockchain.

3) Retrieving Data

The DU starts the retrieve operation by searching the required information of encrypted data on the blockchain. DU uses the link from metadata to get the encrypted data from CS. DU checks the integrity of encrypted data with hash code obtained from metadata set. DU runs the decryption algorithm with his/her private key SK, the public parameters PK and the ciphertext CT.

If the attribute set S of DU satisfies the access structure A which is embedded in the ciphertext CT, then the algorithm can decrypt the CT and DU get the original message M.

V. SECURITY ANALYSIS

In this section, the security of proposed model is proven through the following cases of adversaries' attempts.

Case 1: Supposing that the data stored by some users in the cloud can be access by the adversaries, however the adversaries cannot alter or modify the data as well as cannot view the stored data. Even if the data

is altered, all the users can know such action. Thus, the data is tamper resistant and can be safely preserved.

Proof: The blockchain actually maintain the hash code of all uploaded data and the characteristics of blockchain ensure that the stored data cannot be modified. The data stored on blockchain cannot be altered or delete once confirmed. If the metadata in the blockchain is wanted to modify by the adversaries, they must try the extensive work to construct a new main chain. Such kind of action is neraly impossible. Moreover, if the data is tampered, it can be verified by comparison with the hash code preserved in the blockchain.

Case 2: Supposing that the adversaries attempt to cover the real content by preserving some piece of metadata which is different from the existing metadata, however, the early metadata which is stored on blockchain always exist and it is more legally effective, thus, the adversaries cannot defraud with the false metadata.

Proof: Generally, a judicial system will be needed to adjudicate with some evidence when two different metadata exist. However, no judicial evidence will be needed in our approach because each metadata has a timestamp that shows the time of preservation in the blockchain.

Case 3: Suppose adversaries can pinpoint all blockchain transactions and read the data. If he/she cannot understand the encrypted data, the adversaries will not be able to steal the sensitive information.

Proof: As all data are encrypted and then stored, and thus no one can see the real contents of the preservation as long as the private key is not compromised.

VI. CONCLUSION

In this paper, the CP-ABE is applied to efficiently control access right to achieve the benefits of blockchain technology for university data. All metadata that represent the university data are saved on an immutable and distributed data storage called blockchain. A user who doesn't have access right will not be able to view the data, and the stored data can verify that it had been modified. Thus, it provides a data integrity, authentication and reliable system. A CP-ABE based fine grained access control of data has also been presented in a university scenario. The university data can be stored at cloud storage and still accessible to only those users whose attribute values satisfy the access policy. The proposed model is part of our ongoing research thus, an experimental study will be conducted on the proposed model and evaluate the empirical result to improve the model in the future.

REFERENCES

[1] Q. Liu, Q. Guan, X. Yang, H. Zhu, G. Green, and S. Yin, "Education-Industry Cooperative System Based on Blockchain," in 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN), Shenzhen, 2018, pp. 207–211, doi: 10.1109/HOTICN.2018.8606036.

[2] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[3] Hyperledger, "Hyperledger-fabricdocs Master documentation." [Online]. Available: http://hyperledger- fabric. readthedocs.io/ en/ release/ prereqs.html.

[4] A. Sahai and B. Waters, "Fuzzy identity-based encryption," Springer, vol. 3494, pp. 457–473, 2005.

[5] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-Policy Attribute-Based Encryption," in 2007 IEEE Symposium on Security and Privacy (SP '07), 2007, pp. 321–334, doi: 10.1109/SP.2007.11.

[6] M. Sharples and J. Domingue, "The Blockchain and Kudos: A Distributed System for Educational Record, Reputation and Reward," in Adaptive and Adaptable Learning, 2016, pp. 490–496.

[7] D. J. Skiba, "The Potential of Blockchain in Education and Health Care," Nursing Education Perspectives, vol. 38, no. 4, p. 220, Aug. 2017, doi: 10.1097/01.NEP.0000000000000190.

[8] A. Grech and A. F. Camilleri, "Blockchain in education," Luxembourg: Publications Office of the European Union 2017, no. 132 S. (JRC Science for Policy Report), pp. 1–125, 2017.

[9] R. Gavriloaie, W. Nejdl, D. Olmedilla, K. E. Seamons, and M. Winslett, "No Registration Needed: How to Use Declarative Policies and Negotiation to Access Sensitive Resources on the Semantic Web," in The Semantic Web: Research and Applications, 2004, pp. 342–356.

[10] Ting Yu and M. Winslett, "A unified scheme for resource protection in automated trust negotiation," in 2003 Symposium on Security and Privacy, 2003., 2003, pp. 110–122, doi: 10.1109/SECPRI.2003.1199331.

[11] J. Li, N. Li, and W. H. Winsborough, "Automated Trust Negotiation Using Cryptographic Credentials," in Proceedings of the 12th ACM Conference on Computer and Communications Security, New York, NY, USA, 2005, pp. 46–57, doi: 10.1145/1102120.1102129.

[12] T. T. Thwin and S. Vasupongayya, "Blockchain-Based Access Control Model to Preserve Privacy for Personal Health Record Systems," Security and Communication Networks, 2019. doi: 10.1155/2019/8315614

[13] T. T. Thwin and S. Vasupongayya, "Blockchain Based Secret-Data Sharing Model for Personal Health Record System," in 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA), 2018, pp. 196–201, doi: 10.1109/ICAICTA.2018.8541296.